

U

Ultraviolet Radiation: Distribution and Variability

JULIAN GRÖBNER

European Ultraviolet Calibration Center,
Physikalisch-Meteorologisches Observatorium Davos,
World Radiation Center, Davos Dorf, Switzerland

Article Outline

Glossary
Definition of the Subject and Its Importance
Introduction
Factors Affecting Surface Ultraviolet Radiation
Future Directions
Bibliography

Glossary

Effective surface albedo The effective surface albedo represents the reflectivity of a homogeneous surface which produces the same downwelling UV irradiance as in the presence of an inhomogeneous surface.

Erythmal UV radiation UV radiation weighted with the nominal action spectrum for the spectral sensitivity of Human skin to sunburn (erythema).

Solar zenith angle Angle between the sun and the normal to the Earth's surface. An SZA of 0° corresponds to an overhead sun (solar elevation 90°), while a SZA of 90° corresponds to the sun at the horizon.

Spectroradiometer Instrument using a grating or prism to split the incoming radiation in its spectral components so as to measure the intensity of the solar spectrum at individual wavelengths.

Stratosphere Part of the atmosphere between approximately 10 and 40 km which is characterized by increasing atmospheric temperatures with height

due to the photo dissociation of ozone molecules by solar UV radiation.

Irradiance Radiation flux through a unit surface area.

Radiation dose Amount of radiation reaching a subject during a defined time interval. A daily dose corresponds to the total radiation reaching the surface during one whole day.

UV Radiation Radiation in the wavelength range 100–400 nm. It is subdivided into three regions called UV-C (100–280 nm), UV-B (280–315 nm), and UV-A (315–400 nm).

Definition of the Subject and Its Importance

Solar UV radiation is the most energetic radiation that is able to penetrate the atmosphere and reach the Earth's surface. Due to the very low intensities of the radiation in this wavelength range, its investigation started comparably late with respect to the other spectral regions of the solar spectrum. Dorno was the first to investigate systematically the seasonal and spectral variability of solar UV radiation at Davos, Switzerland. Starting in 1907, his measurements allowed to determine the distinct differences of radiation in this wavelength range, which were later explained by the absorption of atmospheric ozone (mainly located in the stratosphere). Using a specially designed spectroradiometer, he discovered the relationship between the cut-on of the solar spectrum in the UV-B and the path through the atmosphere. In view of his achievements, UV radiation at the time was also called Dorno radiation. In the 1960s, Paul Bener, also from Davos, continued the investigations initiated by Dorno, by building specially designed spectroradiometers that were unique in his time. Systematic measurements under different atmospheric and environmental conditions provided for the first time a benchmark dataset to use as reference for validating radiative transfer models (RTMs), which were emerging at the time. Thus, the

dependence of solar UV radiation on clouds, aerosols, surface reflectivity, and altitude were investigated systematically. Moreover, polarization measurements of the sky radiance were obtained at that time, and are still among the most precise measurements available. Recently, the interest for UV radiation has been linked to its biological effects on humans and on the biosphere in general. The discovery of decreasing ozone levels due to anthropogenic chemicals released in the atmosphere in the early 1980 s has led to intense research activity in the field of solar UV measurements to determine the effect of declining ozone levels on UV radiation and its implications for the environment.

In this entry, the spatial distribution of solar UV radiation at the surface will be discussed, as well as the most important parameters responsible for variations in intensity of the surface solar UV radiation.

Introduction

Solar ultraviolet radiation falling on the Earth's surface originates from the sun. On its passage through the atmosphere, the radiation is scattered and absorbed depending on a variety of factors. Even though the mechanisms are qualitatively similar to what occurs at longer wavelengths, several important differences exist: Ultraviolet radiation is strongly absorbed by atmospheric ozone, located high up in the atmosphere. Furthermore, ultraviolet radiation is strongly scattered by air molecules.

Ultraviolet radiation is classified into three wavelength regions: Radiation at wavelengths just shorter than visible light is called UV-A and extends from 315 to 400 nm. The radiation at shorter wavelengths is more energetic and is classified as UV-B radiation, from 280 nm to 315 nm. At even shorter wavelengths between 100 and 280 nm, the radiation is termed UV-C and is fully absorbed by atmospheric trace gases before reaching the Earth's surface [1]. Due to the spectral characteristics of the ozone absorption cross sections, UV radiation shows a dramatic decrease in intensity – over several orders of magnitude – between approximately 330 and 290 nm. This short wavelength cutoff at around 290 nm is therefore mainly influenced by ozone absorption.

Solar ultraviolet radiation contains only a small part of the total energy of the solar radiation reaching

the Earth's surface. Nevertheless, radiation at these short wavelengths is energetic enough to affect biological tissue, such as breaking DNA bonds [2]. Ultraviolet radiation has detrimental effects on plants, human beings, and underwater life forms such as plankton, fish, larvae, and algae.

Information on the level of ultraviolet radiation reaching the Earth's surface can be obtained from ground-based measurements, radiative transfer modeling, and more recently can be inferred from measurements by satellite sensors.

The most precise and accurate method is to measure the incoming UV radiation from the surface using spectroradiometers [3]. These instruments measure the intensity of radiation in very narrow wavelength bands and can thereby reconstruct the solar spectrum with high fidelity. Since the UV spectrum shows such important intensity changes with wavelength, the knowledge of the solar spectrum gives the most complete knowledge on the surface UV radiation.

Narrowband multifilter radiometers measure the solar spectrum in several narrow UV wavelength bands with resolutions of the order of 2–10 nm. This type of instrument is more compact and robust than a spectroradiometer, but requires a spectroradiometer for its calibration [4].

The most common type of instrument measuring solar UV radiation is a broadband filter radiometer. This instrument uses filters to be sensitive to a broad wavelength interval, and returns a single value representing the solar UV spectrum weighted with the instrument spectral sensitivity. Various types of such broadband instruments exist, depending on what spectral range they are sensitive: UV-B, UV-A, or the total UV range. While these instruments are easy to operate and require very little maintenance, their calibration is complex and necessitates experienced and well-equipped laboratories [5]. Furthermore, due to the gradual deterioration of their filters, these instruments need frequent recalibrations if long-term measurements are required.

Radiative transfer models (RTMs) solve the radiation equation to retrieve the surface UV radiation from a prescribed atmospheric state [6, 7]. While current RTMs solve the radiation equation with sufficient precision, the results are practically limited by the knowledge of the model parameters that are input to the

model. These model parameters (essentially the composition of the atmosphere and the vertical distribution of their components) are usually not very well known and induce corresponding large uncertainties in the computed UV radiation [8].

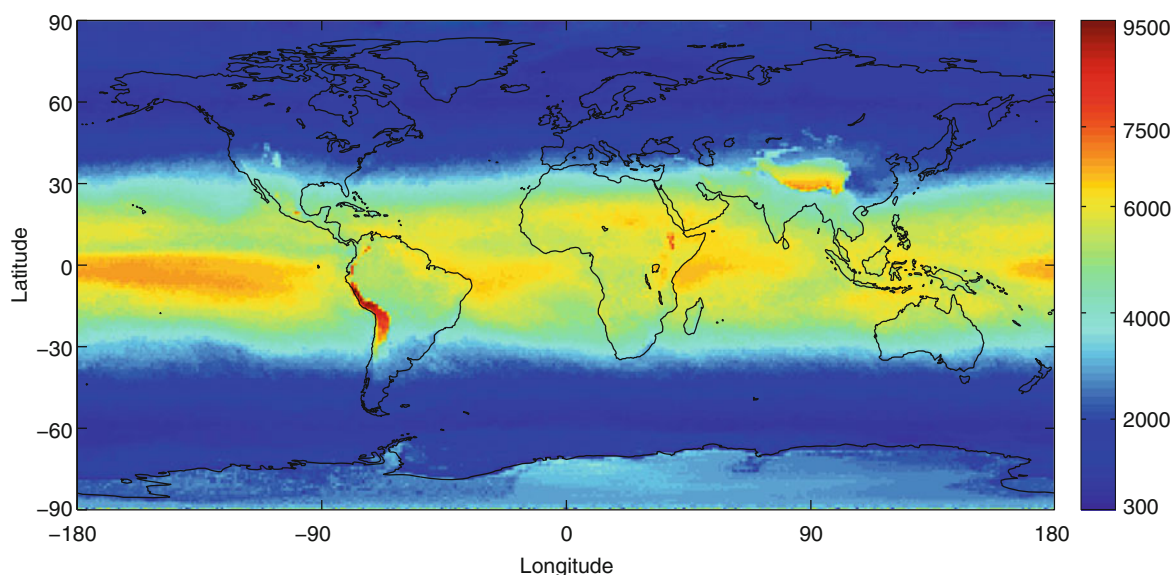
Knowledge of the spatial distribution of UV radiation is readily obtained from remote-sensing instruments located on satellite platforms. These instruments determine UV radiation from the backscattered solar radiation that is reflected by the atmospheric layers and the Earth's surface [9]. The surface UV radiation is then obtained from elaborate model calculations. Due to the difficulty inherent in such an approach, the accuracy of the retrieved UV radiation, although determined over large areas with high spatial resolution, has much larger uncertainties than the corresponding surface point-like measurements. For example, a typical problem faced by satellite sensors is the distinction between snow and clouds, which both have a high reflectivity as seen from space. In the first case, surface UV radiation is enhanced, whereas in the second case it is strongly reduced, depending on cloud type and thickness [10]. Additionally, comparisons between surface UV measurements

and different satellite sensors have shown that the radiation absorbed by tropospheric aerosols is not well taken into account by satellite sensors and retrieval algorithms. Thus, UV radiation in polluted areas can be overestimated by up to 30% [11, 12].

Factors Affecting Surface Ultraviolet Radiation

The map shown in Fig. 1 displays the daily erythemal UV dose reaching the Earth's surface, averaged over the whole year 2005. The map was obtained from measurements by the Ozone Monitoring Instrument (OMI) [13]. Each pixel has a size of $1 \times 1^\circ$. As can be seen in the figure, the highest UV radiation is found close to the equator, where the solar elevation is the highest. In addition, the Andes and the Himalayas mountain ranges are clearly visible due to the much higher UV radiation with respect to the surrounding areas. Less obvious is the reduced UV radiation over large parts of the Amazonian rainforest, due to the predominant cloud cover over this area compared to the Atlantic and Pacific Ocean at the same latitude.

In what follows, the parameters affecting surface UV radiation will be described in detail.



Ultraviolet Radiation: Distribution and Variability. Figure 1

Average daily dose of surface solar erythemal UV radiation in J/m² from the OMI Satellite for the year 2005

Earth to Sun Distance

Due to the point-like size of the sun as seen from Earth, the radiation intensity coming from it decreases as the inverse square of the distance between it and the Earth. This variability is of geometric nature and is therefore independent of wavelength, that is, radiation at all wavelengths is affected equally. Since the Earth path around the Sun follows an ellipse in which the sun is at one focus, slight changes in the distance during one annual rotation are evident. The Earth is closest to the sun in early January and farthest from it in early July. The relative difference between the minimum and maximum intensity is 6.8%. This radiation variability has consequences on the geographical distribution of UV radiation since the maximum occurs during the winter in the Northern hemisphere and correspondingly in (austral) summer in the Southern hemisphere and vice versa. Thus, purely from geometrical considerations, the radiation intensity is higher in summer in the Southern hemisphere than in the Northern hemisphere.

Solar Zenith Angle

The largest variation of surface UV radiation is due to the elevation of the sun above the horizon. The dependence of solar UV radiation on solar zenith angle (SZA) is responsible for the diurnal variation of cloud-free solar UV radiation, which normally has a maximum at local noon. The solar zenith angle is also responsible for the latitudinal dependence of surface UV radiation, as shown in Fig. 1. The overall latitudinal gradient dominates the UV variability on a global scale. This pronounced variability is much higher than at longer wavelengths and is due to the absorption of UV radiation by atmospheric ozone, as explained below.

When the solar zenith angle is low, that is, the sun is high in the sky, the path through the atmosphere is short, so that both the absorption and scattering processes are minimized with respect to high SZA when the path through the atmosphere is much larger. This dependence on SZA is mainly due to the atmospheric ozone contained in the atmosphere, since it strongly absorbs UV radiation on its passage through the atmosphere. As is shown in Fig. 2, the diurnal shape of UV-B radiation is much steeper compared with the variation at longer wavelengths, which are not affected by atmospheric ozone.

Clouds

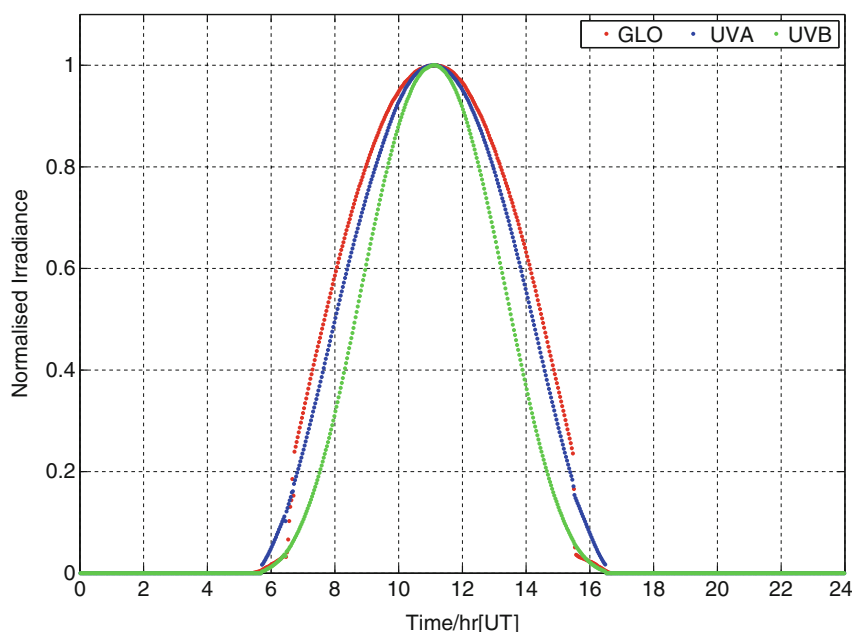
After SZA, clouds represent the most important factor responsible for surface UV variability [14]. Due to their high temporal and spatial variability, clouds are the main reason why it is so difficult to accurately predict UV radiation. Surface UV radiation can be attenuated by up to 99% under certain specific thick cloud types, such as cumulonimbus clouds. In the presence of scattered cumulus clouds, UV radiation is frequently enhanced by the lensing effect from tall adjacent vertical cloud sides [15]. This enhancement can be up to 20% and occurs usually only during short periods of time for particular solar and cloud situations. Figure 3 displays the UV radiation measured at Ispra, Italy between 1992 and 2004. The red curve represents the radiation level possible on a cloud-free day, while the individual points are actual measurements. The large scatter is predominantly due to attenuation by clouds, while the overall yearly variability is due to seasonal changes in solar elevation.

Molecular Scattering

Air molecules in the atmosphere scatter incoming radiation according to its wavelength. As shown by Lord Rayleigh in 1871, the scattering efficiency scales with the inverse fourth power of wavelength ($1/\lambda^4$), thereby scattering short wavelengths much more than longer ones [16]. This relationship, for example, leads to the blue color of the sky in the visible radiation spectrum. For UV wavelengths, the effect is even stronger and is the reason why a large fraction of the incoming solar UV radiation is scattered and reaches the Earth as diffuse radiation. Indeed, even at low SZA, as much as half of the surface UV radiation flux is scattered solar radiation.

Ozone

The inverse correlation between UV Radiation and atmospheric ozone is well documented, and has been thoroughly discussed in the popular and scientific literature. As has been shown, a 1% decrease in total atmospheric ozone leads to an approximately 1% increase in UV-B radiation [17]. The relationship between UV radiation and ozone is strongly dependent on wavelength because the efficiency of ozone in



Ultraviolet Radiation: Distribution and Variability. Figure 2

Typical diurnal variation of global solar radiation (0.3–3 μm), red curve, UV-A radiation, blue curve, and UV-B radiation, green curve, during a clear day at Davos. The discontinuity in global radiation around both 7 and 16 UT is due to surrounding mountains blocking the path of the direct radiation from the sun

absorbing UV radiation increases exponentially with decreasing wavelength.

Due to this inverse relationship, a global decrease in total column ozone would raise UV levels to a level that would become deleterious to many life forms living on the Earth.

The inverse correlation between UV radiation and ozone is generally masked by the much larger variability due to SZA and clouds. In Fig. 4, the correlation between UV-B radiation and ozone is shown by normalizing the irradiance at 305 nm by that at 324 nm. The resulting ratio remains very sensitive to ozone, while becoming insensitive to parameters such as clouds or aerosols, which both affect radiation at 305 and 324 nm nearly equally.

Stratospheric Ozone Hole

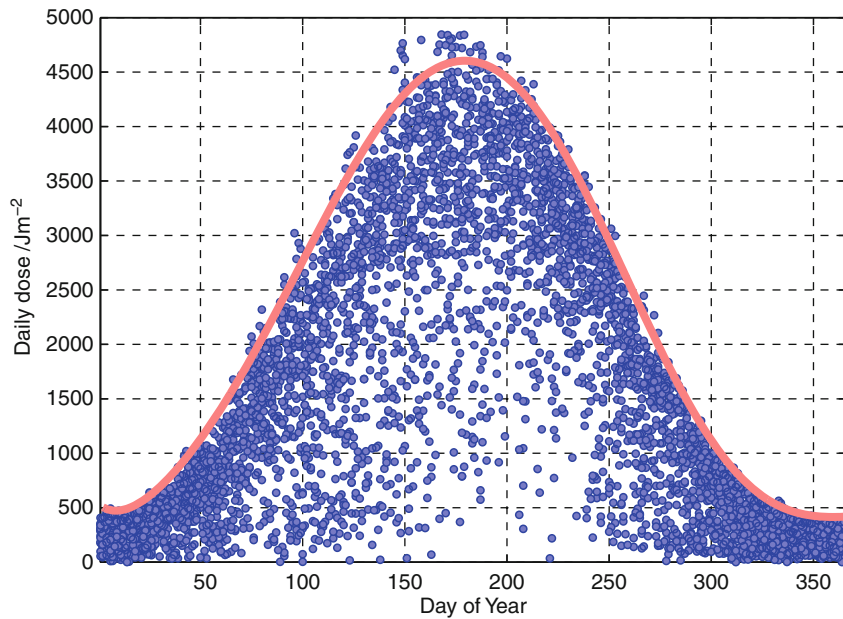
A dramatic ozone decrease has been occurring in the spring period over the Antarctic continent since the early 1980s [18]. Man-made chemical species accumulate in the winter period over the Antarctic, due to atmospheric conditions specific to the Antarctic

continent during winter. In the presence of sunlight, these chemical species react with ozone molecules in the stratosphere and destroy a large fraction of the atmospheric ozone. As much as two thirds of the total ozone contained in the atmosphere can thus be destroyed, which leads in turn to important enhancements in the UV radiation reaching the Earth's surface.

Due to the dynamical nature of the polar vortex, ozone-poor air occasionally reaches also Punta Arenas and Ushuaia at the southern extreme of the South American continent. For instance, during periods of significant ozone reductions, biologically effective UV radiation at Punta Arenas in Chile were more than a factor of 2 higher than during normal conditions [19].

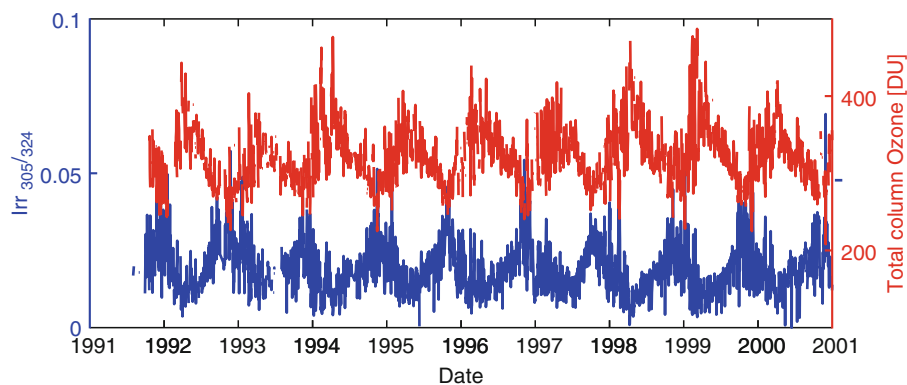
Aerosols

Aerosols in the atmosphere both scatter and absorb radiation. The amount and nature of these processes depend directly on the microphysical properties of the aerosols. Due to the large spatial and temporal variability in aerosols created by anthropogenic or natural



Ultraviolet Radiation: Distribution and Variability. Figure 3

Daily erythemally weighted UV radiation dose at Ispra, Italy for the years 1992–2004. Each point represents the radiation dose reaching the Earth's surface during a single day. The red curve represents the maximum radiation dose expected on a cloud-free day. The large scatter around this curve is due to the effect of clouds, which can reduce UV radiation by up to 99%. Instances of UV radiation higher than the theoretical cloud-free maximum are due to enhanced UV radiation caused by the lensing (or funnel) effect from the sides of cumulus clouds. This effect can locally and temporarily increase UV radiation by up to 20%



Ultraviolet Radiation: Distribution and Variability. Figure 4

Measurements of UV radiation and total column ozone at Ispra, Italy from 1991 to 2001. The UV radiation at 305 nm is strongly affected by ozone, whereas radiation at 324 nm is only weakly attenuated by ozone. The ratio between radiation at 305 nm and 324 nm (*blue curve*) is largely insensitive to clouds and other parameters that are only weakly depending on wavelength, thus highlighting the dependence on ozone. The total column ozone (*red curve*) measured at the same site shows a systematic yearly shape that is anti-correlated with the UV radiation at 305 nm

processes, the quantification of these processes remains very challenging.

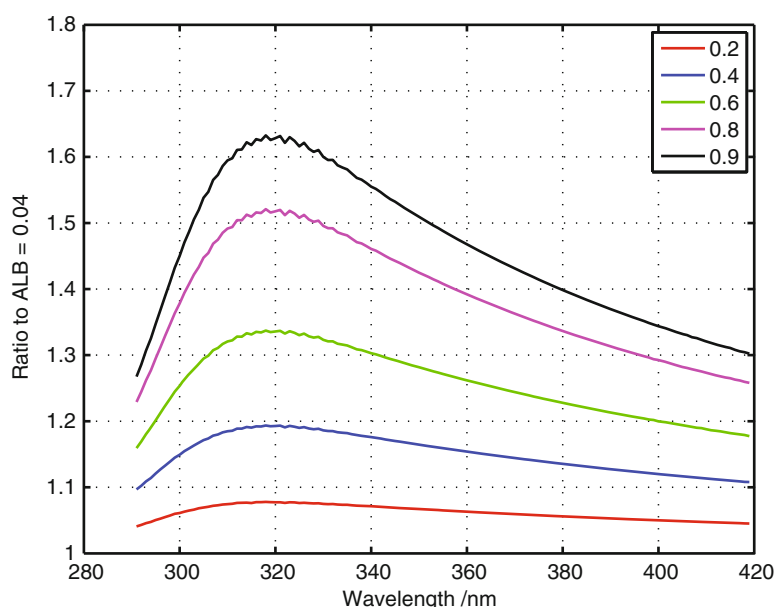
Two main parameters are used to quantify the effect of aerosols on UV radiation. The amount of aerosols, termed the aerosol optical depth (AOD) is a measure of the total aerosol load of the atmosphere. Incoming direct solar UV radiation is attenuated due to scattering and absorbing processes, which are a direct function of AOD. The relative efficiency of aerosols in either scattering or absorbing radiation is described by a second parameter, called the single-scattering albedo (SSA). This parameter can be obtained from microphysical properties of the aerosols. In practice, SSA is very difficult to quantify for a particular atmospheric situation due to the large variability of the aerosol composition. UV radiation can be attenuated by 30% or more in presence of heavy aerosol loads [20].

Surface Reflectivity

Radiation reflected from the surface is backscattered by the atmosphere and thus increases diffuse radiation. This enhancement of diffuse radiation occurs from the scattering that occurs on the molecules and aerosols

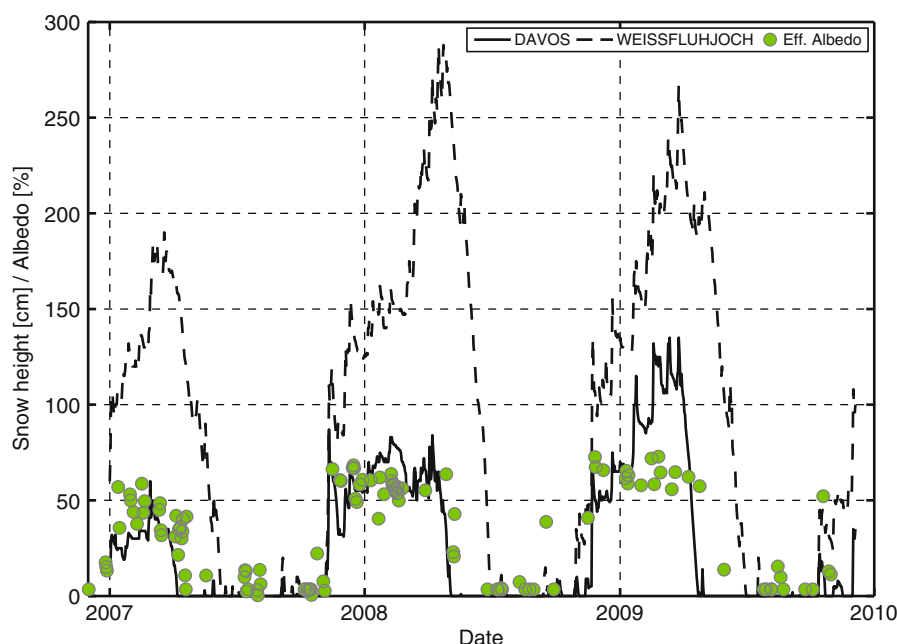
present in the lower atmospheric layers. Due to the large molecular scattering effect in the UV wavelength range, a significant fraction of the reflected radiation is scattered back to the Earth surface, thereby enhancing the downwelling solar UV radiation. As can be seen in Fig. 5, this enhancement increases toward shorter wavelengths as is expected from the increased molecular scattering efficiency.

The largest enhancement occurs in presence of snow, which has a reflectivity of up to 90% [21]. Studies have shown that in regions with a complete snow cover UV radiation can be enhanced by up to 50% comparatively to a snow-free situation [22]. Radiative transfer models are generally used to infer the effect of surface albedo on UV radiation. As is shown in Fig. 4, the enhancement due to surface albedo has a specific spectral signature, which is mainly due to the interplay between two parameters: At wavelengths above ≈ 330 nm, where ozone absorption is negligible, the enhancement is due to Rayleigh scattering. Around 320 nm, a distinct maximum is found followed by a gradual decrease toward shorter wavelengths, which is due to the increasing effectiveness of tropospheric



Ultraviolet Radiation: Distribution and Variability. Figure 5

Enhanced UV radiation for different surface albedo values with respect to the case with a UV albedo of 0.04. The gradual increase with decreasing wavelength is due to the increasing scattering efficiency of air molecules, while the decrease below about 320 nm is caused by tropospheric ozone absorption



Ultraviolet Radiation: Distribution and Variability. Figure 6

Effective surface albedo in percent, as determined from UV measurements during cloud-free periods at Davos (elevation 1,590 m). The black and dotted lines represent the actual snow height measured in Davos and Weissfluhjoch (elevation 2,540 m, horizontal distance from Davos, 2 km), respectively. During winter, the effective albedo is as high as 70%, enhancing UV radiation by up to 30% with respect to snow-free conditions. During summer, the effective albedo drops to values close to 2–3%

ozone in absorbing UV radiation. Since the reflected radiation is scattered several times before reaching the surface, the likelihood of being absorbed by tropospheric ozone is enhanced.

The enhancement of surface UV radiation due to reflected radiation from the surface that is backscattered to the surface is influenced by the surrounding area around the measurement site. Indeed, model calculations demonstrate that radiation reflected as far away as 30 km can influence the locally incident radiation [23, 24]. State-of-the art three-dimensional radiative transfer models can calculate the actual radiation for an arbitrary surface environment composed of a complex patchwork of albedo conditions, and be only limited by the spatial resolution of the model. In contrast, most common radiative transfer models can only prescribe a single albedo value for the whole surface, which is then called an effective surface albedo [25]. This effective surface albedo represents the overall reflection characteristics of a usually

inhomogeneous surface over a circular area with a radius of up to 30 km.

Figure 6 shows the effective albedo determined for Davos from 2007 to 2009 by combining measurements and radiative transfer calculations. As can be seen in the figure, during the winter period when Davos and the surrounding areas are fully covered by snow, the effective albedo has an average value of about 0.6, whereas during the summer months the surface albedo decreases to values around 0.03. The corresponding radiation increase with respect to the snow-free summer period is as large as 30%. It should be noted however that, due to the lower solar elevation in winter than summer, the daily radiation dose is much higher in summer than winter.

Future Directions

The intensity of solar UV radiation in the future will depend on possible changes in several factors, in

relation to global and regional changes in the climate system. These factors include ozone, aerosols, clouds, and surface reflectivity, which are all expected to change on both a regional and global scale due to climate change. Changes in stratospheric ozone can be modeled quite accurately using global chemistry models. Their results indicate that the total ozone levels will not only recover as a consequence of the measures taken under the Montreal protocol and its amendments, but that a “super-recovery” of ozone is likely to occur around 2050 [26]. This super-recovery will manifest itself with higher ozone levels than prior to the global ozone decrease, with correspondingly reduced UV levels, depending on the region considered [27].

Bibliography

Primary Literature

- Coblentz WW (1932) The Copenhagen meeting of the second international congress on light. *Science* 76:412–417
- Diffey BL (1991) Solar ultraviolet radiation effects on biological systems. *Phys Med Biol* 36(3):299–328
- Gröbner J (2001) Characterisation of spectrophotometers used for spectral solar ultraviolet radiation measurements. *Rad Prot Dosimetry* 97:415–418
- Dahlback A (1996) Measurements of biologically effective UV doses, total ozone abundances, and cloud effects with multichannel, moderate bandwidth filter instruments. *Appl Opt* 35:6514–6521
- Hülser G (2007) Characterization and calibration of ultraviolet broadband radiometers measuring erythemally weighted irradiance. *Appl Opt* 46:5877–5886
- Chandrasekhar S (1960) Radiative transfer. Dover Press, New York. ISBN 0-486-60590-6
- Stamnes K et al (1988) Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media. *Appl Opt* 27:2502–2509
- Weihs P et al (1997) Accuracy of spectral UV model calculations 1. Consideration of uncertainties in input parameters. *J Geophys Res* 102(D1):1541–1550
- Herman J et al (1999) Distribution of UV radiation at the Earth's surface from TOMS-measured UV-backscattered radiances. *J Geophys Res* 104(D10):12059–12076
- Arola A et al (2003) A new approach to estimating the albedo for snow-covered surfaces in the satellite UV method. *J Geophys Res* 108:4531. doi:10.1029/2003JD003492
- McKenzie R et al (2001) Satellite retrievals of erythemal UV dose compared with ground-based measurements at northern and southern midlatitudes. *J Geophys Res* 106:D20. doi:10.1029/2001JD000545
- Arola A et al (2005) Assessment of TOMS UV bias due to absorbing aerosols. *J Geophys Res* 110:D23211. doi:10.1029/2005JD005913
- Tanskanen A et al (2007) Validation of daily erythemal doses from Ozone Monitoring Instrument with ground-based UV measurement data. *J Geophys Res* 112:D24544. doi:10.1029/2007JD008830
- Bener P (1964) Investigation on the influence of clouds on the ultraviolet sky radiation. Contract AF 61(052)-618, Tech. Note 3, Davos-Platz, Switzerland
- Mims FM III, Frederick JE (1994) Cumulus clouds and UV-B. *Nature* 371:291
- Rayleigh (1871) On the scattering of light by small particles. *Phil Mag* 41, 107, 274
- Madronich S et al (1998) Changes in biologically active ultraviolet radiation reaching the Earth's surface. *J Photochem Photobiol B Biol* 46(1–3):5–19
- Farman JC et al (1985) Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction. *Nature* 315:207–210
- Kirchhoff V et al (1997) UV-B enhancements at Punta Arenas, Chile. *J Photochem Photobiol B Biol* 38:174–177
- Kylling A et al (1998) Effect of aerosols on solar UV irradiances during the photochemical activity and solar ultraviolet radiation campaign. *J Geophys Res* 103:26051–26060
- Blumthaler M et al (1988) Solar UVB albedo of various surfaces. *Photochem Photobiol* 48(1):85–88
- Bernhard G et al (2007) Ultraviolet and visible radiation at Barrow, Alaska: climatology and influencing factors on the basis of version 2 National Science Foundation network data. *J Geophys Res* 112:D09101. doi:10.1029/2006JD007865
- Degünther M et al (1998) Case study on the influence of inhomogeneous surface albedo on UV irradiance. *Geophys Res Lett* 25:3587–3590
- Mayer B, Degünther M (2000) Comment on “Measurements of erythemal irradiance near Davis station, Antarctica: effect of inhomogeneous surface Albedo”. *Geophys Res Lett* 27: 3480–3490
- Kylling A et al (2000) Determination of an effective spectral surface albedo from ground-based global and direct UV irradiance measurements. *J Geophys Res* 105(D4):4949–4959
- Austin J et al (2006) Ensemble simulation of the decline and recovery of stratospheric ozone. *J Geophys Res* 111:D16314. doi:10.1029/2005JD006907
- WMO (2007) Scientific assessment of ozone depletion: 2006. Global ozone research and monitoring project report No. 50. World Meteorological Organization, Geneva

Books and Reviews

- Seckmeyer et al (2001) Instruments to measure solar ultraviolet radiation, part 1: spectral instruments (WMO TD No. 1066), available at <http://www.wmo.ch/pages/prog/arep/gaw/gaw-reports.html>
- Seckmeyer et al., Instruments to measure solar ultraviolet radiation – Part 2: Broadband instruments measuring erythemally

weighted solar irradiance (WMO TD No. 1289), 55 p, July 2008, available at <http://www.wmo.ch/pages/prog/arep/gaw/gaw-reports.html>

Webb et al (2003) Quality assurance in monitoring solar ultraviolet radiation: The state of the art (WMO TD No. 1180), available at <http://www.wmo.ch/pages/prog/arep/gaw/gaw-reports.html>

Zerefos C (1997) Solar ultraviolet radiation: modelling, measurements and effects (NATO ASI Series/Global Environmental Change). ISBN 3-540-62711-1

Underground Cable Systems

YAMILLE DEL VALLE, NIGEL HAMPTON, JOSH PERKEL,
CARYN RILEY
National Electric Energy Test Research and
Application Centre (NEETRAC), Forest Park,
GA, USA

Article Outline

Glossary
Definition of the Subject
Introduction
Cable System Structure
MV, HV, and EHV Cable Systems
Uses of Cables
AC and DC Transmission
Cable Types
Components of the Cable
Cable Manufacture
Failure Processes
Dry Aging Thermoelectric Aging
Future Directions
Understanding Life Expectancy of Cable Systems
Improving Cable Performance with Respect to Aging
Recyclable/Recoverable Cable Designs
Increased Use of Long Length Links
Bibliography

Glossary

Breakdown Permanent failure through insulation.
Cable system Cable with installed accessories.
CCV Catenary continuous vulcanization extrusion line for extruded dielectrics.

Cross-linked polyethylene (XLPE) A thermoset unfilled polymer used as electrical insulation in cables.

Diagnostic test A field test made during the operating life of a cable system. It is intended to determine the presence, likelihood of future failure and, for some tests, locate degraded regions that may cause future cable and accessory failure.

Dielectric loss An assessment of the electric energy lost per cycle. A poorly performing cable system tends to lose more energy per AC cycle. Measurements can be made for selected voltages or over a period of time at a fixed voltage. The stability of the loss, the variation with voltage and absolute loss are used to estimate the condition. Data can be derived from time-based (if sufficient time is taken) or frequency-based test methods.

Electrical trees Permanent dendritic growths, consisting of nonsolid or carbonized micro-channels, that can occur at stress enhancements such as protrusions, contaminants, voids, or water trees subjected to electrical stress. The insulation is damaged irreversibly at the site of an electrical tree.

Ethylene propylene rubber (EPR) A type of thermoset-filled polymer used as electrical insulation in cables and accessories. There are several different formulations of EPR and they have different characteristics. For purposes here, the term also encompasses ethylene propylene diene monomer rubber (EPDM).

Extra high voltage (EHV) Cable systems within the voltage range 161–500 kV, though more often between 220 kV and 345 kV. Also referred to as transmission class, though usually has higher design stress levels than HV.

Extruded dielectrics Insulation such as EPR, HMWPE, PE, WTRXLPE, XLPE, etc., applied using an extrusion process.

Filled insulation Extruded insulations where a filler (carbon black or clay) has been incorporated to modify the inherent properties of the base polymer. This class includes all types of EPR, Vulkene, etc.

High pressure fluid filled (pipe type or HPOF) paper insulated and installed in trefoil in steel pressure pipes and impregnated with high-pressure nondegradable fluid which is maintained at high

pressures by pumping plants – common in USA at HV & EHV.

High voltage (HV) Cable systems within the voltage range from 46–161 kV, though more often between 66 kV and 138 kV. Also referred to as transmission class, though usually has lower design stress levels than EHV.

Jacket An extruded outer polymeric covering for cables designed to protect the cable core and the metallic shielding (wires, tapes, or foils).

Joint A device to join two or more sections of power cable together. A joint includes a connector to secure the cable conductor and a stress controlling/insulating body to manage the electrical stress.

Laminated dielectrics Insulation formed in layers typically from tapes of either cellulose paper or polypropylene or a combination of the two. Examples are the PILC (paper-insulated lead-covered) and MIND (mass-impregnated non-draining) cable designs.

Mass impregnated non-draining Cable (MIND)
A cable design using paper insulation impregnated with a thick compound such that the compound does not leak out when the lead is breached.

Maximum electrical stress The highest level of stress also corresponds to the highest probability of instantaneous failure or, equivalently, the highest rate of electrical aging.

MDCV Mitsubishi Dainichi continuous vulcanization, often called long land die, extrusion line for XLPE, typically for HV & EHV.

Mean or average electrical stress This is most important if the most serious defects are uniformly located throughout the bulk of the insulation.

Medium voltage (MV) Cable systems within the voltage range from 6 kV to 46 kV, though more frequently between 15 kV and 35 kV. Also referred to as distribution class.

Metallic shield A concentric neutral surrounding the cable core. The shield provides (to some degree) mechanical protection, a current return path, and, in some cases, a hermetic seal (essential for impregnated cables).

Minimum electrical stress This is most important if cable system reliability is determined by the performance of accessories or if the electrical design or installation method of accessories degrades cable performance.

Paper insulated lead covered (PILC) A cable design using paper insulation impregnated with a fluid and encased in lead to prevent the fluid from leaking out of the insulation.

Partial discharge A low voltage (mV or μ V) signal resulting from the breakdown of gas enclosed in a dielectric cavity. The signals travel down the cable system and may be detected at the end, thereby enabling location.

PE-based Extruded insulations that do not have an incorporated filler (carbon black or clay). This class includes all types of HMWPE, PE, WTRXLPE, XLPE, etc.

Polyethylene (PE) A polymer used as electrical insulation in cables.

Power frequency A substantially sinusoidal waveform of constant amplitude with an alternating frequency in the range of 49–61 Hz.

Self-contained fluid filled (FF or LPOF or PILC)
Paper or paper polypropylene laminated (PPL) insulated with individual metal sheaths and impregnated with a dielectric fluid. Where used these are common in land cable applications. This type of cable is one of the first to be installed in the 1890s.

Shielded cable A cable in which an insulated conductor is encapsulated in a conducting “cylinder” that is connected to ground.

Space charge Quasi-permanent injected charge that is trapped within the insulation of a cable system. This charge is sufficient to modify the applied AC and impulse voltage stresses.

Splice A joint.

Tan δ (TD) The tangent of the phase angle between the voltage waveform and the resulting current waveform.

Termination A device that manages the electric stress at the end of a cable circuit, while sealing the cable from the external environment and providing a means to access the cable conductor. Devices referred to as elbows or potheads are types of terminations.

VCV Vertical continuous vulcanization extrusion line for XLPE, typically for HV & EHV

Water tree retardant cross-linked polyethylene (WTRXLPE) A thermoset polymer used as electrical insulation in cables that is designed to retard water tree growth.

Water trees Dendritic pattern of electro-oxidation that can occur at stress enhancements such as protrusions, contaminants, or voids in polymeric materials subjected to electrical stress and moisture. Within the water tree the insulation is degraded due to chemical modification in the presence of moisture.

Definition of the Subject

Underground cables have been used from the earliest time as integral parts of the power distribution and transmission system. Compared to their overhead analogues, they have been long regarded as the most critical of components due in part to their high total installed cost, their unique ampacity requirements, and the complexity of their installation. Thus, even from the very first technical papers, the sustainability, through reliability and longevity, has been of paramount importance. This contribution addresses the focus on sustainability today while maintaining a linkage to the lessons that have been already learned.

Introduction

Almost all electric power utilities distribute a portion of the electric energy they sell via underground cable systems. Collectively, these systems form a vast, interlinked, and valuable infrastructure. Estimates for the USA indicate that underground cables represent 15–20% of installed distribution system capacity. This percentage is much closer to 50% in Europe. These cable systems consist of many thousands of miles of cable and hundreds of thousands of accessories installed under city streets, suburban developments, and the countryside. Utilities have a long history of using underground system with some of these cable systems installed as early as the 1890s. Very large quantities of cable circuits were installed in the 1950s–1980s. Today, the size of that infrastructure continues to increase rapidly as the majority of newly installed electric distribution lines are placed underground. There are a number of drivers and brakes that control this process and they are shown in [Table 1](#).

Cable systems are designed to have a long life with high reliability. However, the useful life is not infinite. These systems age and ultimately reach the end of their reliable service lives. Estimates set the design life of

Underground Cable Systems. Table 1 Factors encouraging and discouraging the installation of underground cable systems

Encouraging installation of underground cable systems	Discouraging installation of underground cable systems
Public perception of risk	Capital cost – higher for cable systems
Reliability (frequency of outages) – frequency is lower for underground cable systems	Return to service after failure (repair time) – longer repair time
Total operating cost – lower electrical losses & lower repair and maintenance costs	Need for reactive compensation on long lengths AC systems
Public perception of visual impact – lower impact	

underground cable systems to be in the range of 30–40 years. Today, a large portion of this cable system infrastructure is reaching the end of its design life, and there is evidence that some of this infrastructure is reaching the end of its reliable service life. This is a result of natural aging phenomena as well as the fact that the immature technology used in some early cable systems is decidedly inferior compared to technologies used today. Increasing failure rates of these older systems are now adversely impacting system reliability, and it is readily apparent that action is necessary to manage the consequences of this trend.

Cable System Structure

The name “cables” is given to long current-carrying devices that carry their own insulation and present an earthed outer surface [1]. In this context, overhead lines, for example, are not considered as cables. Power cables have a coaxial structure: Essentially, they comprise a central current-carrying conductor at line voltage, an insulation surrounding the conductor, and an outer conductor at earth potential. AC cables are generally installed as a 3-phase system, and hence, the outer conductor should only carry fault and loss currents. In practice, a more sophisticated construction is

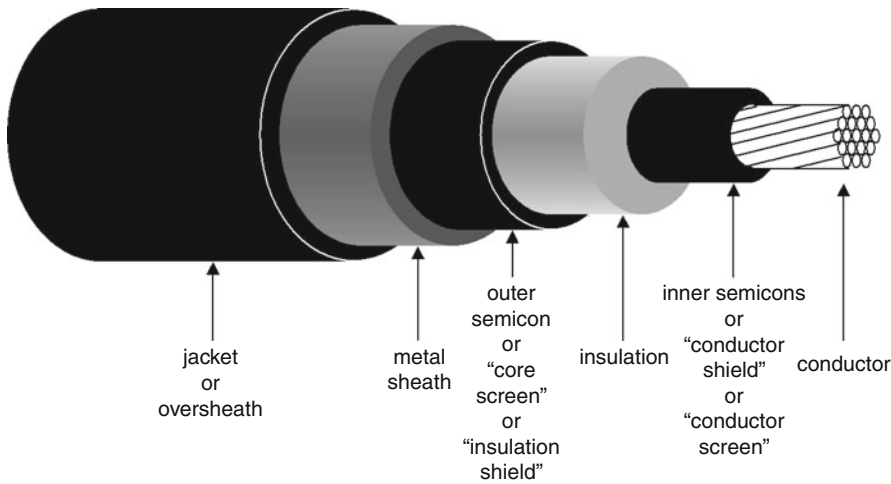
adopted. The interfaces between the metal conductors and the insulation (laminated or polymeric) would tend to include contaminants, protrusions, and voids, features that would lead to electrical stress enhancement, accelerated aging, and premature failure [2, 3]. To overcome this, a “semicon” layer, a conductive paper or polymeric composite, is placed at both interfaces. The inner semicon, the insulation, and the outer semicon ensure the interfaces are smooth and contaminant-free. Surrounding this cable are layers to protect the cable during installation/operation and carry the loss/fault currents. These layers also serve to keep out water, which may lead to water treeing in polymer insulations or elevated losses in laminated insulations. A schematic diagram of a power cable is shown in Fig. 1.

MV, HV, and EHV Cable Systems

Cable systems that are used for distribution and transmission purposes are generally categorized according to the voltage rating:

- Medium voltage (MV): 6–36 kV
- High voltage (HV): 36–161 kV
- Extra high voltage (EHV): 161–500 kV (or more)

There is no international consistency on the distinction between “distribution” and “transmission.” Which means that although clear in the bulk of the ranges, the edges/transitions get somewhat blurred. Table 2 shows the relationship between the voltages and voltage classes of cable systems for selected regions of the world.



Underground Cable Systems. Figure 1
Cut-away section of a typical power cable

Underground Cable Systems. Table 2 Voltage classes for AC cable systems in selected regions

	MV	HV	EHV
Asia	6.6, 10, 15, 22, 33	66, 90, 110, 132, 154, 161	187, 220, 275, 345, 400, 500, 525
Europe	11, 20, 33	63, 66, 90, 110, 132, 145, 150	220, 275, 380, 400
North America	15, 25, 35, 46	69, 115, 138	220, 230, 345, 500
	Distribution	Transmission	
Nordic	20	90, 132, 145	220, 400

Electrical Stresses

The operation of a cable system is very dependent upon the electrical stresses (E) to which it is subjected, for example:

- Dielectric heating $\propto E^2$
- Probability of failure $\propto 1 - \exp(E/\alpha)^\beta$
- Insulation aging $\propto E^n$

Thus, to understand the many issues associated with sustainability, it requires a firm understanding of the electric stresses in the whole system.

Alternating Current (AC) The electrical stress within a cable is given by [1].

$$E = \frac{V}{x \ln\left(\frac{R}{r}\right)} \quad (1)$$

where V is the applied voltage, r is the radius over the inner semiconductive screen, R is the diameter over the insulation, and x is the intermediate radius (between r and R) at which the electric stress is to be determined.

The probability of failure depends upon the electrical stress and increases with stress. The effect of an increased stress is most commonly estimated using the Weibull probability function [4]:

$$P_f = 1 - \exp\left\{-\left(\frac{E}{\alpha}\right)^\beta\right\} \quad (2)$$

where P_f is the cumulative probability of failure, i.e., the probability that the cable would have failed if the stress is increased to a value E . The two parameters, α and β , are known, respectively, as the characteristic stress and the shape parameter.

Inspection of Eq. 1 shows that the electric stress varies with the position within the cable. There are three potentially useful stresses that can be considered:

- Maximum stress at conductor screen
- Mean geometric average stress for the whole insulation
- Minimum stress at the core screen

The decision as to which of these to consider is an important one and is guided by the potential modes of failure:

- Maximum: The highest level of stress also corresponds to the highest probability of instantaneous failure or, equivalently, the highest rate of electrical aging. This is most important if the most serious cable defects are located on or near the conductor screen.
- Mean: This is most important if the most serious defects are uniformly located throughout the bulk of the insulation.
- Minimum: This is most important if cable system reliability is determined by the performance of accessories or if the electrical design or installation method of accessories degrades cable performance. It is also important if the most serious cable defects are located on or near the core screen.

The range of electrical stresses employed is shown in Table 3.

Direct Current (DC) The stress in AC cables is determined by the capacitance (permittivity) of the structure. However, in DC cable systems, the resistance (resistivity) determines the stress [1, 5].

Underground Cable Systems. Table 3 Average stress levels for selected insulations

Type	Stress (kV/mm)	EHV	HV	MV
Fluid filled	Average at core screen	10.0	8	8
	Average at conductor screen	17	14	10.0
EPR	Average at core screen	4	3	2
	Average at conductor screen	8	5	3
XLPE	Average at core screen	5.0	3	2
	Average at conductor screen	11	6	3

The permittivity of insulations is essentially stress and temperature independent (within the range of normal use), whereas the resistivity has significant voltage and temperature dependence.

Inspection (Fig. 2) shows that in the temperature range (20–60°C), the resistivity can change by 2–3

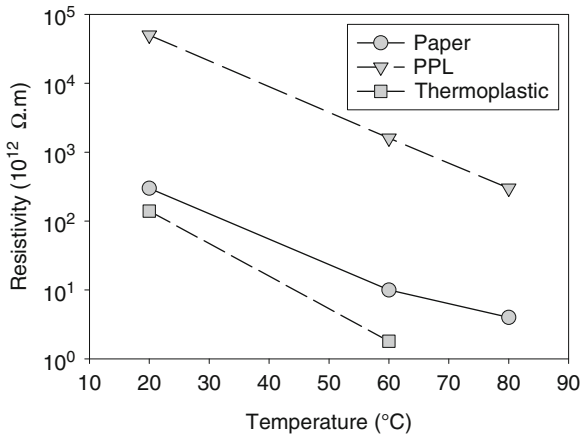
orders of magnitude. The dependence of the resistivity ρ on stress E and temperature T can be characterized equally well by using the formats of Eqs. 3 or 4.

$$\rho = \rho_0 \exp(-\alpha T - \beta E) \quad (3)$$

$$\rho = \rho_0 \exp(-\alpha T) E^{-\gamma} \quad (4)$$

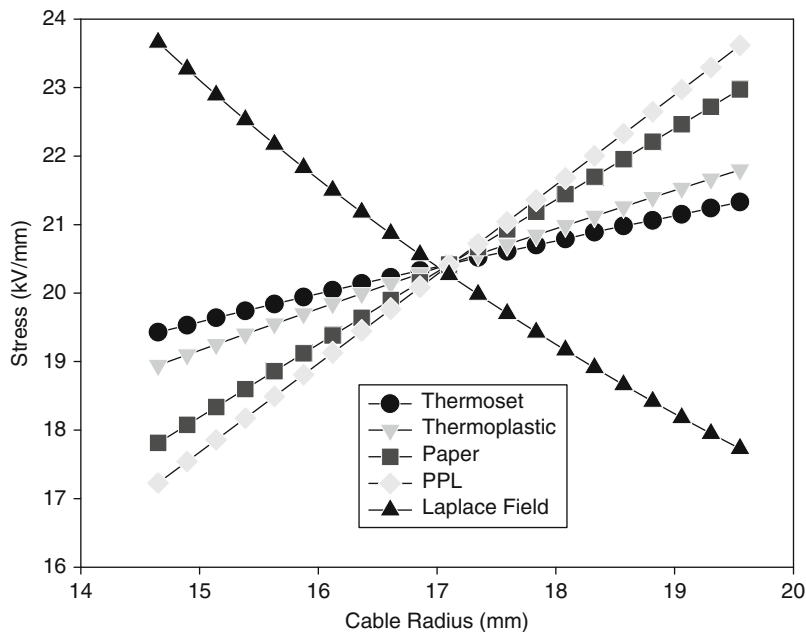
where ρ_0 is the resistivity at reference temperature, α is the temperature coefficient, and β and γ are the stress coefficients of the insulation.

The practical consequence is that the stresses in a DC cable system are primarily determined by the dimensions, voltage, and the temperature distribution. Thus, the design of DC systems is technologically much more challenging than AC. The most obvious manifestation is the difference (with respect to AC systems) of the location of the maximum stress: In AC systems, this is always located close to the conductor (termed Laplace field in Fig. 3); in DC systems, this can move from the conductor (no load case) to the outside (highly loaded case) – this is termed stress/temperature inversion.



Underground Cable Systems. Figure 2

Resistivity of selected insulations at different temperatures



Underground Cable Systems. Figure 3

Electrical stress distribution, under load, for a 100 kV DC cable manufactured using different insulants. The power loss W is the same for all designs

One of the ways to calculate the electrical stress distribution E_r across the cable insulation at radius r can be shown below:

$$E_r = \frac{\delta V (r/r_2)^{\delta-1}}{r_2 \left[1 - (r_1/r_2)^\delta \right]} \quad (5)$$

$$\delta = \frac{\alpha W / 2\pi\lambda + \beta V / (r_2 - r_1)}{\beta V / (r_2 - r_1) + 1} \quad (6)$$

where λ is the thermal conductivity, r_1 is the conductor screen radius, r_2 is the radius over insulation, $W = I^2 R$ is the conductor loss, and V is the applied voltage. The α and β parameters are extracted from the resistivity data using Eq. 3.

Uses of Cables

Power cables are commonly used in underground or underwater (submarine) connections. Cables are placed at strategic points of the transmission grid to supplement overhead lines or, in some cases, they can form the whole “backbone.” Interconnections between networks are particularly well suited to cable solutions [3] for security of supply reasons.

Cables may also be used in other applications rather than just underground or underwater. For example, overhead covered conductors allow smaller phase clearance between the conductors on medium-voltage overhead lines. Objects, particularly tree branches, may touch the lines without tripping or customer outage. This has led to substantial improvements in service reliability (e.g., [5]). In most cases, the aluminum alloy conductor is covered with black UV-resistant cross-linked polyethylene and filled with grease, to provide corrosion protection and longitudinal watertightness. Arcing guides are applied at insulator tops, to protect the line from arcing damage. Another application is the “Powerformer” [6] and the related “motorformer.” These new generators are able to supply electricity directly to the high voltage grid without the need for a step-up transformer. It is suitable for power generation at output voltages of several 100 kV. One example of this is the Troll motorformer project on a natural gas rig in the North Sea. The new concept is based on circular conductors for the stator winding,

and it is implemented by using proven high-voltage cable technology. Thus, the upper limit for the output voltage from the generator is only set by that of the cable.

AC and DC Transmission

Cable systems are used in both alternating current (AC) and direct current (DC) schemes. The cable designs used in each case are outwardly very similar and have many identical design elements. However, the detailed engineering and the materials used are very different. AC is the globally preferred means of transferring electric power. This form of transfer makes it straightforward to generate electricity and to transform voltages up and down. This means of transfer accounts for more than 98% of the global power infrastructure.

In long-distance transmission schemes, especially those that are of interest in the future such as windfarms or solar plants, there are very significant advantages in using DC over AC. System instabilities caused by connecting regions with slightly different AC phases and frequencies are obviated. Capacitive charging current implies that there is a maximum useful length of AC cables without the use of shunt reactors. DC cables are therefore particularly useful for long-distance submarine connections. Furthermore, the lack of electromagnetic effects under DC conditions eliminates the skin effect in which the conductor resistance can rise by up to 20% at 50 Hz. The drawbacks for a DC solution are that the terminal equipment for AC/DC conversion is seen as more costly and less efficient than AC transformers. (Reviews on AC/DC power transmission are contained in [7].) Thus, the system must be sufficiently long to be economically viable. In addition, the control of the electrical stress in a DC cable insulation is also much more difficult due to thermally induced stress inversions and stress modifications due to trapped space charge. The space charge issues are particularly onerous at interfaces, such as those that occur at accessories (joints and terminations).

Before 2000, cables using lapped/impregnated (oil or compound) insulations were almost exclusively used for DC transmission up to 500 kV and very long distance. Even today this technology is preferred for the highest voltages (EHVDC) and distances. However, the recent development in the HVDC and lower power

Underground Cable Systems. Table 4 HVDC projects using extruded cable systems

Project	Type of Project	Power MVA	Voltage rating KV	Cable length km	In service Year
Gotland/SE	S	60	80	140	1999
Directlink/AU	TL	180	84	390	2000
Tjæreborg/DK	WIND	8	10	8	2000
Cross Sound/US	S	330	150	82	2002
Troll/NO	OFF	80	60	70	2004
Estlink/FI	S	350	150	105	2006
Valhalla/NO	OFF	78	150	292	2009
NordEon/DE	WIND	400	150	390	2009
Transbay/US	S	400	200	95	2010

S Subsea Interconnection, OFF Offshore Power Supply, WIND Windpower Delivery to Shore, TL Trading Link

range has seen the symbiotic development of cable systems and converter technologies. The pace of this development has been quite rapid such that the current installed lengths (not at similar voltages) of paper and extruded cables are approximately equal at 22,500 km. A selection of HVDC projects using extruded cables is shown in Table 4.

Cable Types

There are, in general, four types of underground power cable technologies in use today:

- *Polymeric*: Cross-linked polyethylene (XLPE), water tree retardant cross-linked polyethylene (WTRXLPE), or ethylene propylene rubber (EPR).
- *Self-contained fluid filled (FF or LPOF or PILC)*: Paper or paper polypropylene laminated (PPL) insulated with individual metal sheaths and impregnated with a dielectric fluid. Where used these are common in land cable applications. This type of cable is one of the first to be installed in the 1890s.
- *Mass impregnated non-draining (MIND or solid)*: Paper insulated with individual metal sheaths and impregnated with an extremely high viscosity polybutene compound that does not flow at working temperatures – common at MV and submarine HVDC.
- *High pressure fluid filled (Pipe Type or HPOF)*: Paper insulated and installed in trefoil in steel pressure pipes

and impregnated with high pressure nondegradable fluid which is maintained at high pressures by pumping plants – common in USA at HV & EHV.

The typical electrical stresses employed in cables are shown in Table 3.

Up until the mid-1980s, paper-insulated cables (PILC, MIND & LPOF) were the system of choice at medium and high voltages. However, improvements in polymeric cables and accessories plus environmental concerns with the impregnation (dielectric fluid and the associated lead sheaths) have led to a significant reduction in the use of paper cables for land applications. In particular, at MV, there has been a strong preference for XLPE & WTRXLPE cables over EPR (except in Italy and USA). XLPE/WTRXLPE and EPR have emerged as the favored polymeric insulations through the 90°C continuous operating temperature that can be achieved when they are used. This temperature matches that which can be attained when fluid-filled lapped insulations (paper and PPL) are used. In contrast, LDPE and HDPE are limited to operating temperatures of 70°C and 80°C, respectively. As a consequence, these insulations have fallen into disuse.

Tables 5 and 6 identify some of the main advantages of the respective technologies at distribution and transmission voltages.

The situation today is:

- Almost all new HV and EHV systems that are being installed are new build/expansion.

Underground Cable Systems. Table 5 Advantages (+) and disadvantages (–) of MV cable insulations

XLPE 11–46 kV	EPR 11–46 kV	WTRXLPE 11–46 kV	Paper 11 to 46 kV (PILC and MIND technologies)
+ No risk of oil leakage	+ No risk of oil leakage	+ No risk of oil leakage	– Oil leakage in PILC technology 0 Reduced compound leakage in MIND technology
+ Very low dielectric losses	0 Medium dielectric losses	+ Low dielectric losses	– High dielectric losses
+ Simple accessory designs which require competent workmanship	+ Simple accessory designs which require competent workmanship	+ Simple accessory designs which require competent workmanship	0 Simple/robust accessories which require high level of workmanship – installation skills disappearing with aging of the workforce
– Likely to suffer from water trees if no metal sheath	0 Less likely to suffer from water trees if no metal sheath	0 Less likely to suffer from water trees if no metal sheath	+ Metal sheath required to contain fluid, thus water is excluded
0 Standard Size	0 Standard Size Reduced size available in special designs	0 Standard Size	+ Small size – higher design stresses – important for installation in ducts

Underground Cable Systems. Table 6 Advantages (+) and disadvantages (–) of HV & EHV cable insulations

XLPE 66–500 kV	EPR 66–138 kV	Paper 66–500 kV (low-pressure and high-pressure technologies)
+ No risk of oil leakage	+ No risk of oil leakage	0 Oil leakage is a concern
+ Simple design	+ Simple design	– Complicated design – High maintenance burden
+ Very low dielectric losses	0 Medium dielectric losses	0 Medium dielectric losses
+ Acceptable fire performance	+ Acceptable fire performance	– Fire performance in tunnel and substation applications is a concern due to the copious supply of flammable oil
+ Simple accessory designs which require high level of workmanship	+ Simple accessory designs which require high level of workmanship	0 Simple accessories which require high level of workmanship – installation skills disappearing with aging of the workforce
0 Growing track record	– Limited track record	+ Extensive track record
– Likely to suffer from water trees if no metal sheath	– Likely to suffer from water trees if no metal sheath	+ Metal sheath required to contain fluid, thus water is excluded
– Large size	– Large size	+ Small size – higher design stresses

- The majority of HV cables, *already installed* within the existing system, are insulated with paper (83% paper and 17% polymeric).
- There is very little replacement of existing paper cables: Most of installed capacity is based on paper.
- Most HV transmission is by overhead lines (OHLs).
- HV cables are replacing OHLs in environmentally sensitive areas but there is limited impact (globally) on new OHLs.

- XLPE cables make up 80–90% of the HV cable capacity *presently being installed*.

The reasons that there has been little replacement of paper cables are:

- Paper cables continue to operate reliably.
- Operating temperatures have been considerably below the temperature limits
- The designs of XLPE cables are too large to fit existing rights of way or pipes that have been designed for paper cables. This is of special importance in the USA where there are many paper cables and small tight ducts. In this area, the key task is to develop polymeric cables designs that are flexible, small (i.e., working at high stresses), and easy to join as it is essential to use the existing pipes/ducts. Present designs address these issues, and now, cables exist that match the size and performance of paper cables [8, 9].

Components of the Cable

Conductor

Conductors in the USA tend to be based on the American wire gauge (AWG), but, in the rest of the world, are based on IEC228 and are therefore described using the metric convention. Stranded conductors have generally comprised concentric layers, but, in order to make them smoother and more compact, they are often specially shaped in rolling mills nowadays. When operating at high voltages and currents, the AC current is preferentially carried more in the outer than in the inner conductors (the skin effect). In addition, the electromagnetic fields induce eddy currents (proximity effect). These effects tend to increase the conductor resistance under AC above that which is seen under DC. The AC/DC ratio ($R_{AC/DC}$) can be as large as 1.15. This increased resistance serves to increase the Joule heating losses within the conductor and increases the temperature of the cable. Thus, special “Milliken” conductors may be used for large conductor designs to reduce the AC/DC ratio.

Conductors are virtually all made from either copper or aluminum. Copper has the advantage of being more conductive, and therefore, requires less material to carry a given current. Copper conductors, therefore, have the advantage of being small. However, the cost of

aluminum, although variable, is lower than that of copper, and even though it has a lower conductivity, it is often used as the conductor. An additional advantage of aluminum over copper is that the conductors are lower in weight even though more volume of material is used. At MV, aluminum is preferred whereas, at HV and EHV, the smaller size of copper provides the greatest advantage.

Semicon

Semiconductive screening materials are based on carbon black, manufactured by the complete and controlled combustion of hydrocarbons. The carbon conducting medium is dispersed within a paper or polymer matrix depending upon the type of cable involved. In early designs of polymeric cable, paper tapes or painted carbon screens were used. These displayed extremely poor service performance as they were seen to dramatically accelerate the growth of water trees. In these cases, failures occurred many years short of the anticipated design life. Current designs of polymeric cables (XLPE, WTRXLPE, or EPR) use extruded polymeric materials to construct the semiconducting screen. Optimal performance is obtained when the screens are coextruded with the insulation in a closed “true triple” extrusion head.

The concentration of carbon black, in both paper and polymeric screens, needs to be sufficiently high to ensure an adequate and consistent conductivity. The incorporation must be optimized to provide a smooth interface between the conducting and insulating portions of the cable. The smooth surface is important as it decreases the occurrence of regions of high electrical stress [10]. To provide the correct balance of these properties, it is essential that both the carbon black and polymer matrix be well engineered.

The same care needs to be paid to the manufacture of the matrix materials (paper or polymer) for semicons as for the insulation. In the case of extruded screens, the chemical nature of the polymers is subtly different from those used for the insulations because of the need to incorporate the carbon black. The carbon black and other essential additives (excluding cross-linking package) are compounded into the matrix. The conveying and compounding machinery used are designed to maintain the structure of the carbon black within

a homogeneous mix. Before the addition of the cross-linking package, filtration may be applied to further assure the smoothness of the material, to a higher standard than that provided by the compounding process.

The smoothness of the extruded cable screens is assured by extruding a sample of the complete material in the form of the tape. The tape is optically examined for the presence of pips or protrusions. Once detected, the height and width of these features are estimated, thereby enabling width-segregated concentrations to be determined. When using such a system, care needs to be exercised when examining the present generation of extremely smooth (low feature concentration) screens, as the area of tape examined needs to match the likely number of detected features: Smooth screens require larger areas of examination.

Insulation

It seems clear that, globally, XLPE or its WTRXLPE variants are the most commonly used cable insulations currently being installed. Insulating XLPE compounds need to fulfill a number of requirements. They should act as thermoplastic materials within the extruder and crosshead. They should cross-link efficiently with the application of high temperatures and pressures within the vulcanization tube. They should be immune to thermal degradation throughout the cable manufacture process and operation at the maximum cable temperature for the life of the system. They must display an extremely low occurrence of the features that can enhance the applied electrical stress and thereby lead to premature failure. To deliver these requirements, it is essential that the greatest care is paid to the design and manufacture of the polymer and the engineering of the appropriate cross-linking and stabilizing packages.

The manufacturing technology employed for XLPE compounds to be used for power applications needs to ensure the highest level of cleanliness at all points of the production chain. The sequence comprises three main parts: base polymer manufacture, addition of a stabilizing package, and addition of the cross-linking package. The most common route for cross-linking in cables (XLPE, WTRXLPE, and EPR) is peroxide cure – thermal degradation of an organic peroxide after extrusion causes the formation of cross-links between the molten polymer chains.

Cleanliness The cleanliness of XLPE and WTRXLPE insulation materials may be assessed by converting a representative sample of the polymer into a transparent tape and then establishing the concentration of any inhomogeneities. This is not possible for monosil as the chemistry occurs immediately prior to the cable extrusion. The inhomogeneities are detected by identifying variations in the transmission of light through the tape. To gain the required level of consistency and sensitivity, the tape is inspected by an automated optical system.

Metal Sheath

For many years, lead or lead alloys were the main materials used for the metal sheath layer. This is principally because the low melting temperature allows the lead to be extruded at a temperature of approximately 200°C over the polymeric cable. The main disadvantages of lead are its high density (11,400 kgm⁻³) leading to a heavy product; environmental concerns; and the tendency to creep, flow, or embrittle under cyclic temperature loadings. This latter effect has led to a number of cases where the sheath ruptured. There are also environmental concerns relating to the use of lead. Its use may become restricted by European Union (EU) directives. In 2000, the EU Commission officially adopted the waste electrical and electronic equipment (WEEE) and reduction of hazardous substances (ROHS) proposals. The ROHS proposals required substitution of lead and various other heavy metals from 2008. To address these problems a number of materials have been used.

- *Extruded aluminum*: This has excellent mechanical performance but requires a large bending radius and can be difficult to manufacture since it requires corrugations. It can be heavy.
- *Aluminum foil*: This is light and easy to manufacture but small thicknesses (0.2–0.5 mm) do not give mechanical protection; the strength comes from the polymer oversheath. It relies on adhesive to make a watertight seal, and it can suffer from corrosion.
- *Copper foil*: This is light, easy to manufacture but not as flexible as aluminum foil. It also relies on adhesive to make a watertight seal.
- *Welded copper*: This is strong, robust, and capable of carrying significant current but difficult to

manufacture since it requires corrugation and it is difficult to ensure perfect longitudinal welds in practical situations.

- *Welded stainless steel*: This is strong, robust, and capable of carrying significant current but has similar manufacturing difficulties to welded copper.

Oversheath (Jacket)

In addition to the meticulous attention that must be paid to the insulation system, care also needs to be taken with the oversheath layer. One of the earliest lessons that was learned from MIND and PILC cables (where a lead sheath is employed) is that installation damage and corrosion can compromise the metallic sheath [11, 12]. Once sheath integrity is lost, water may enter and impregnant leak out. In these cases failures in service occurred quite rapidly. These issues were soon resolved on neoprene tape or extruded oversheaths were used. When extruded cables were introduced in the mid-1960s, it was believed that there was no longer a need for an oversheath as the insulating polymers were themselves waterproof. However, the phenomena of water treeing soon showed that a good-quality cable jacket was extremely valuable even at the low electrical stresses prevalent at MV.

The vast majority of HV and EHV XLPE cables are of the “dry design” type, which means that a metal barrier is included. The purpose of the metal barrier is to protect the core within from mechanical damage, carry fault and loss currents, and to exclude water from the construction. (The electrical aging rate is significantly higher in the presence of moisture.) The metal barrier is a key part of the cable design and much care needs to be taken as this significantly affects how a cable system may be installed in practice. In a similar way to that at MV, the metal layer is itself protected by a polymeric oversheath. Due to the critical performance needed from the oversheath, there are a number of properties that are required: good abrasion resistance, good processing, good barrier properties, and good stress crack resistance. Experience has shown that the material with the best composite performance is an oversheath that is based on polyethylene.

Cable Manufacture

Stages of Cable Manufacture

The stages in cable manufacture may be summarized as:

1. *Conductor manufacture*: This involves
 - *Wire drawing* to reduce the diameter to that required
 - *Stranding* in which many wire strands and tapes are assembled
 - *Laying up*: the assembly of noncircular (Milliken) segments into a quasi-circular construction
2. *Core manufacture*: This involves Paper (LPOF, MIND, PILC) cables
 - *Lapping* in which the core of the cable is formed by wrapping semiconducting and insulation tapes around the conductor with prescribed tensions and paper tape overlaps
 - *Laying up (triplexing)* for three core cables
 - *Drying* which improves the dielectric properties of the cable core and is carried out at elevated temperatures and under vacuum
 Polymer (XLPE, WTRXLPE, and EPR) cables
 - *Triple extrusion* in which the core of the cable is formed comprising the inner semicon, insulation, and outer semicon.
 - *Cross-linking* which is carried out directly after extrusion (peroxide cure).
 - *Degassing* in which peroxide cross-linking by-products are removed by heating offline. The diffusion time depends upon temperature and insulation thickness.
3. *Cable manufacture*: This involves Paper (LPOF, MIND, PILC) cables
 - *Impregnation* where the dry paper tapes are impregnated at elevated temperatures with dry/degassed dielectric fluid
 - *Metal sheathing*: the application of a metal impregnant enclosure
 - *Oversheathing*: the application of high strength extruded polymeric oversheath (jacket)
 - *Armouring*: the application of high strength metal components (steel) to protect the cables; essential for submarine cables
 - *Routine testing*: voltage withstand and ionization factor tests

Polymer (XLPE, WTRXLPE, and EPR) cables

- *Core taping* during which cushioning, protection, and water exclusion layers inner semicon, insulation, and outer semicon are applied over the extruded core
- *Metal sheathing*: the application of a metal moisture and protection layer
- *Oversheathing*: the application of high strength extruded polymeric oversheath (jacket)
- *Armouring*: the application of high strength metal components (steel) to protect the cables; essential for submarine cables
- *Routine testing*: voltage withstand and partial discharge tests

Methods of Core Manufacture for Extruded Cables

All of the production processes will be common to all methods of manufacture, with the exception of the extrusion process where there are three types of peroxide cross-linking methods. (Cross-linking is often referred to as vulcanization or curing.)

- VCV: Vertical continuous vulcanization
- CCV: Catenary continuous vulcanization
- MDCV: Mitsubishi Dainichi continuous vulcanization, often called long land die

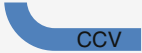
In all of these processes, the three layers of the cable core are extruded around the conductor. This un-cross-linked core then passes directly into the curing tube; this is where differences in the processes become apparent. In the moisture cure approach, which takes place offline after extrusion, the manufacturing process is considerably simplified as the length of the tube following extrusion only has to be long enough for the thermoplastic core to cool sufficiently to prevent distortion.

The general relationship between manufacturing method, the insulation technologies, and voltage classes is shown in [Table 7](#).

Failure Processes

Power cable systems are designed to be high-reliability products. The failure of a major power cable is likely to have a considerable effect on the power transmission grid and may take several days/weeks to repair. If it is under the sea, it may take months/years to repair and cost well in excess of \$3 million (€3 million). Cables have a good service history. The majority of cable failures are caused by external influences such as road diggers or ship anchors. Cable system failures may also occur at joints and terminations.

Underground Cable Systems. Table 7 General correlation of extrusion equipment with extruded insulation materials

	MV			HV			EHV		
	EPR	WTR XLPE	XLPE	EPR	WTR XLPE	XLPE	EPR	WTR XLPE	XLPE
Catenary vulcanization 	X	X	X	X	–	X	–	–	X
Vertical vulcanization 	–	–	–	–	–	X	–	–	X
Horizontal vulcanization 	–	–	–	–	–	X	–	–	X

Underground Cable Systems. Table 8 System performance for a 20kV XLPE ([13])

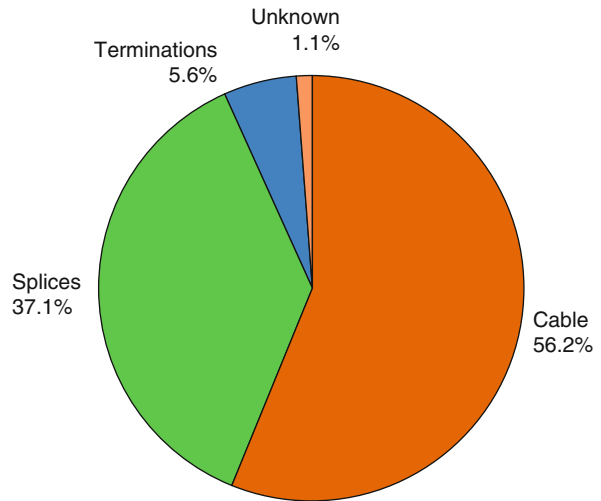
Class	Fault rate (#/year/100 cct km)
Total system	2.0
Third party damage	1.0
Accessories	0.9
Cable	0.1

A study of MV cable systems in France [13] has shown that the failure rate for paper cables is 3.5 failures per 100 cable km per year which compares with a rate of 2 failures per 100 cable km per year for XLPE cables. The failure rate for XLPE has been classified in Table 8. This analysis shows that the XLPE cables form the most reliable component of the cable system. Further inspection would suggest that the XLPE cable system is 2.5 times more reliable than the paper system when third party damage is excluded. (It is assumed that a paper cable is as likely to get dug up as an XLPE one!) This estimate may well overstate the case and illustrates one of the major problems of evaluating field failure statistics, namely, the differing ages of the populations. XLPE data are based on a population which is 15 years old whereas the paper data are based on a population that is 30–40 years old; thus, higher failure rates of the paper system are expected since it contains older devices; this would be quite independent of the intrinsic reliabilities of the systems.

The service performance of cable system accessories (joints and terminations) depends upon six essential elements:

- Connector design
- Joint/termination body design
- Installation methods
- Installation quality
- Operating conditions

A more recent study in the USA [11] estimated a median failure rate of 3.5 failures per 100 cct miles per year. Figure 4 shows an estimate of how these failures are disbursed by the sources of failure. At first sight, these figures seem to be quite different to the EDF study; however, it should be recognized that there will



Underground Cable Systems. Figure 4

Estimated dispersion of North American MV cable system failures by equipment type

be considerable variation due to utility and geographical location. For example, the average percentage of failures ascribed to splices is 37%; however, this can range from 5% to 80% depending upon utility. Nevertheless, these studies both show that it is important to consider the cable system as a whole entity.

In general, the causes of insulation breakdown will include the causes listed below. Table 9 shows how the types of defects may be related to the failure modes for the different voltage classes of cables. Many of these defects are shown schematically in Fig. 5.

- Extrinsic defects (contaminants, protrusions, or voids) caused during manufacture or installation. These would normally lead to electrical treeing or direct breakdown soon after production of the void and cable energization.
- Water treeing (“wet aging”) caused by water leakage through the sheath or inappropriate design or deployment of a cable without a water barrier. Water trees lead to a weakening of the insulation and electrical treeing or direct breakdown.
- Water ingress caused by failure of the metal water barriers. Water increases the dielectric loss and thereby leads to local overheating. Failure proceeds by thermal runaway in a region with compromised local breakdown strength.

Underground Cable Systems. Table 9 Role, in general terms, of defects in cable failure modes

Cause of failure	MV	HV	EHV
Extrinsic defects <i>Contaminants</i> <i>Delaminations</i> <i>Protrusions</i> <i>Shield interruptions</i> <i>Voids/cracks</i>	Accelerants for water treeing	Major source of failure Accelerants for electrical treeing	Major source of failure
Water treeing <i>Bowtie</i> <i>Vented – partial</i> <i>Vented bridging</i>	Major source of failure for extruded	Can be an issue if robust metallic water barriers are not used	Rarely an issue due to the use of robust metallic water barriers
Water ingress	Major source of failure for laminar	Major source of failure for laminar	Rarely an issue due to the use of robust metallic water barriers
Thermoelectric aging	Rarely an issue due to the low temperatures and stresses of operation	Major source of failure accelerated by extrinsic defects	Major source of failure accelerated by extrinsic defects

- Thermoelectric aging: The combination of the electric field, acting synergistically with a raised temperature, causes the insulation to weaken over time and for breakdown to occur eventually. This process may not always be significant within the lifetime of a well-designed cable.

Extrinsic Defects

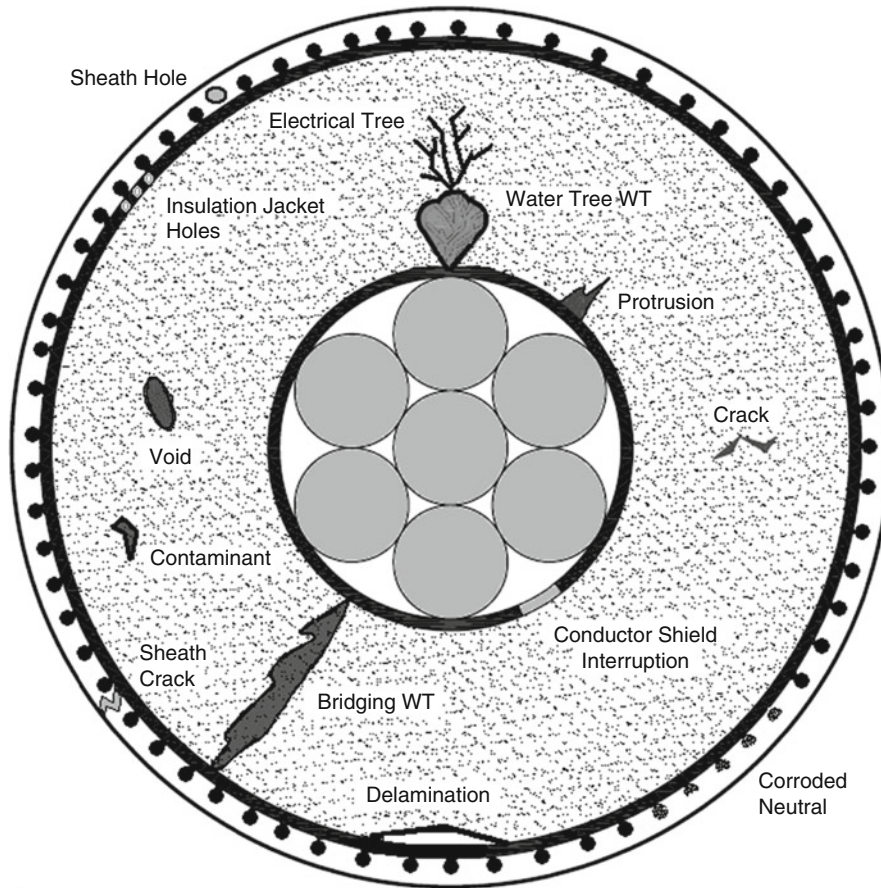
The efforts made have already been described, from the earliest times, by cable manufacturers to exclude and detect contaminants, protrusions, and voids (CPVs) in their products. Contaminants within the bulk of the insulation and protrusions into the insulation from the semicon cause field intensifications that lead to premature failure of the polymer.

The degree of aging follows the empirical Inverse Power Law relationship with the electric stress ($\propto E^n$) [2, 3, 12, 14, 15].

Contaminants and Protrusions Many studies [2, 3, 5, 11, 12, 14–18] have shown the degradation caused by large metallic contaminants. Generally contaminants and protrusion type defects reduce the characteristic strengths of insulators; furthermore, the increasing size of contaminants changes the statistical nature of the failures, making them less scattered or more certain.

The effects may be explained by the fact that metallic contaminants increase the electric stress within their immediate locality such that the local electric stress is higher than the breakdown strength of the insulator. This effect is best described in terms of a stress enhancement factor that acts as a multiplier for the Laplacian stress (i.e., the geometrically calculated stress). It is interesting to note that it is possible to get large stress enhancements at sharp metallic contaminants but that the magnitude of the enhancement falls dramatically with distance from the tip; for a 5- μm radius, the field falls by 50% within 1.5 radii of the tip [10]. Thus, the calculated stress enhancements should be viewed as providing the upper limits of any assessment and it is quite challenging to relate them directly to the likelihood of failure.

The electrical stress enhancements are not only based on the size and concentration but they have a significant influence from the nature (conducting, insulating, high permittivity), the shape (sharp or blunt), and the way that they are incorporated into the matrix. The effect of the shape of contaminants has been assessed [18] using XLPE cups with a Rogowski profile. It was shown that contaminants with irregular surfaces reduced the AC ramp breakdown strength by a greater degree than those with smooth surfaces.



Underground Cable Systems. Figure 5
Typical power cable defects

The local stress enhancement experienced within an insulator will have contributions from the size of the contaminants, their concentration, and the nature (conducting or high permittivity) of the contaminants. This is shown in Eq. 7:

$$\eta = 1 - \frac{1}{\alpha} \left(0.5 \ln \frac{\lambda + 1}{\lambda - 1} - \frac{\lambda}{\lambda^2 - 1} \right) \quad (7)$$

where:

$$\alpha = 0.5 \ln \frac{\lambda + 1}{\lambda - 1} - \frac{1}{\lambda} + \frac{1}{(k - 1)\lambda(\lambda^2 - 1)}$$

$$k = \frac{\epsilon_2}{\epsilon_1} \quad \lambda = \frac{1}{\sqrt{1 - \frac{r}{a}}}$$

η = stress enhancement factor,
 r = radius of the ellipse,

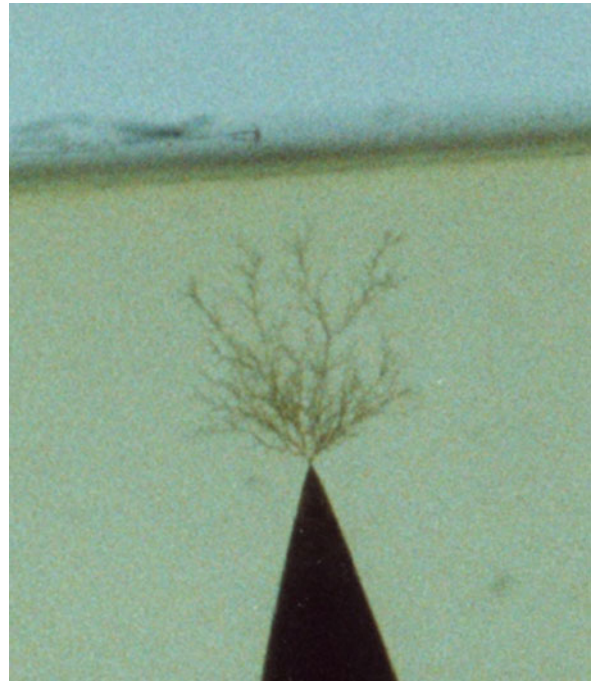
$2a$ = length of the ellipse,
 ϵ_1 = permittivity of the matrix, and
 ϵ_2 = permittivity of the defect.

Voids Voids are likely to lead to breakdown *if* discharges occur inside them. These discharges are known as “partial discharges” since they are not, in themselves, a complete breakdown or “full discharge.” The least problematic (and perhaps the least likely) shaped void is the sphere. The field inside an air-filled, sphere-shaped void is higher than the field in the insulation by a factor equal to the relative permittivity of the solid. For other shaped voids, the field in the void will be higher than this. For example, a void inside XLPE (relative permittivity = 2.3) will have a field inside it of at least 2.3 times that of the field in the XLPE itself. The criterion for a discharge in a void is that the void

field must exceed the threshold described by the Paschen curve (see [12] for example); this is dependent on the size of the void and the gas pressure within it. The Paschen field has a minimum for air at atmospheric pressure for a void diameter of 7.6 μm . The breakdown voltage at this diameter is 327 V. This equates to a void field (ignoring the nonlinear effects) of 43 kV/mm or an applied field of 19 kV/mm within the XLPE. Below this void size, the Paschen field increases rapidly; for example, a 2- μm void would require an applied field of approximately 150 kV/mm to cause discharging. It is clear then that voids of diameter exceeding a few microns are likely to allow severe electrical damage to occur.

Partial Discharges and Electrical Treeing In an electrical discharge, electrons are accelerated by the electric field such that their kinetic energy may exceed several electron-volts. With such energies, collisions with gas molecules may cause further electrons to be released, thus strengthening the discharge or they may cause electroluminescence and the release of energetic photons. The surface of the void is therefore likely to be bombarded by particles (photons or electrons) with sufficient energy to break chemical bonds and weaken the material. In the case of sharp protrusions or contaminants, which give rise to high local electric fields, electrons may be emitted and very quickly acquire the kinetic energy required to cause permanent damage to the insulation. Electrons may accumulate – i.e., they may be trapped – around such defects and cause a further increase in local electric fields. Mechanical stress, which may already be increased due to the modulus and thermal expansion coefficient differences between the host materials and the CPVs, may be further enhanced by electromechanically induced stress. These effects, catalyzed by CPVs in an electric field may lead directly to a breakdown path, but are more likely to lead first to the formation of an electrical tree.

An electrical tree is shown in Fig. 6. For the sake of clarity, this has been grown in a translucent epoxy resin from a needle acting as a protrusion. A breakdown path can be seen to be growing back through the electrical tree from the plane counter-electrode. Electrical trees have a branched channel structure roughly oriented along the field lines. Typically, the diameters of the



Underground Cable Systems. Figure 6
An electrical tree grown in epoxy resin

channels are 1–20 μm , and typically, each channel is 5–25 μm long before it branches. There is considerable evidence that the branching is determined by the local electric field, which is grossly distorted by trapped electrons emanating from the discharges within the tree. There is a considerable body of work on this subject (e.g., [12]). Trees tend to grow more directly across the insulation if they are spindly, so-called branch trees. The other type of tree, the “bush” tree, uses the energy to produce a lot more dense local treeing and, therefore, tends to take longer to bridge the insulation. Because branch trees occur at *lower* voltages, there is a non-monotonic region between branch and bush growth in which failure occurs more quickly at lower voltages. Trees grown from larger voids can be clearly distinguished from those due to protrusions; it is noticeable that the fields must cause the voids to discharge before the trees initiate. After an initial period in which the tree grows fast, the rate of growth decreases. Finally, as the tree approaches the counter-electrode, a rapid runaway process ensues.

Although the electric tree processes are now quite well understood, apparently little can be done to

prevent electrical tree propagation in polymeric insulation once it has started. It is thought that electrical trees take relatively little time to grow and cause breakdown in cables, perhaps a few minutes to a few months. Once an electrical tree has been initiated, the cable can be considered to be “terminally ill.” It is therefore vital to prevent the inclusion of CPVs during cable manufacture and installation. The triple extrusion continuous vulcanization techniques and appropriate cable protection and deployment techniques appear to have been successful in this regard, and cables rarely suffer from these problems. There are more likely to be problems at joints or terminations where high field stresses may inadvertently be introduced or if water trees, described in the following section, are allowed to grow.

Interaction Between Moisture and Temperature

One of the effects of water entering an insulation is that it tends to increase the permittivity and loss of the insulation [1, 19, 20]. The energy lost per cycle within an insulation is proportional to both the permittivity and the loss. Thus, at a given electrical stress, the increasing energy loss due to the presence of water will lead to an elevated temperature. If this temperature becomes too high, other mechanisms come into play and there is a thermal runaway to failure.

Although this happens in principle in all insulations, and may be very important at the high stresses of HV & EHV systems, this is most commonly observed in paper-insulated MV cables. The sensitivity of these cables is due to the fact that the dielectric losses are initially much higher and the paper and oils have a higher propensity to absorb water. Some basic measurements were made by Blodgett [19] who showed that the loss in paper insulations increased with both temperature and moisture. These data can be used within cable system rating models to estimate the thermal equilibrium for selected temperatures and moisture contents. The results are shown in Fig. 7 which shows the increment above the base-case temperature in the form of a contour-plot for selected moisture contents and temperatures. Inspection shows that significant temperature rises can be expected for quite reasonable moisture contents. A separate study has shown that cables removed from service but not failed might be expected to have a median moisture content of 1%, but cables that failed had moistures in excess of

3%. Thus, the thermal runaway mechanism seems to be a reasonable explanation of the observed phenomena. Moreover, the calculations show the difficulty of setting wide-reaching criteria; a common empirically determined maximum permitted moisture level is 3%. However, the impact of this level of moisture is quite different for different cables and temperatures: 3°C, 7.5°C, 14°C for 15 kV/90°C, 35 kV/90°C, and 35 kV/105°C, respectively.

Wet Aging: Water Trees

In the early days of polymer-insulated cables, it was assumed that the polymers would be essentially immune to the deleterious effects of water that were well known in paper cables. Consequently, the first designs of cables were installed with little or no water precautions. Within a few years, a large number of cables started to fail in service. Upon examination, tree-like structures were seen to have grown through the insulation. It was assumed that they continued to grow and failure occurred when the whole insulation was breached. This is the phenomenon of water treeing [1–3, 21].

Many studies have been carried out into the phenomena and its solution. Looking back, it is clear that a number of improvements in cable design, manufacture, and materials have reduced the incidence of cable failures by water treeing. These improvements have included

- Water barriers (metal or polymeric) to exclude the water
- Triple extrusion (all polymer layers extruded at the same time)
- Semiconductive polymer screens to replace carbon paint or paper tapes
- Cleaner insulations
- Smoother semicons
- Internationally recognized approval methods
- Special long life insulations based either on additives or polymer structure

The laboratory studies have concluded that the growth of water trees is affected by:

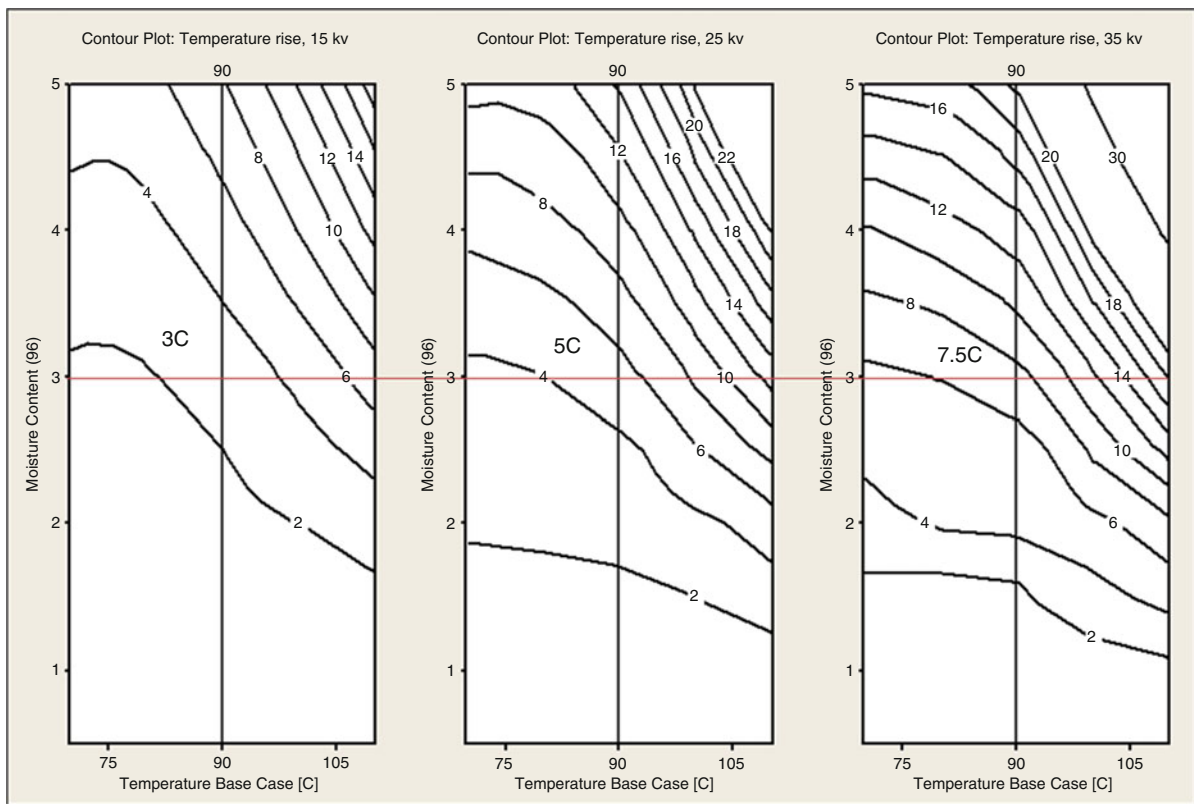
- Test voltage
- Test frequency

- Mean temperature
- Temperature gradient
- Type of material
- Presence of water (external and within the conductor)

There are essentially two types of water trees: (a) vented trees that grow across the insulation and are potentially the most dangerous and (b) bow-tie trees that grow across the insulation and tend to grow to a limiting size without breaching the insulation. These trees do *not* comprise tubules containing water as might be surmised from the earlier description of electrical trees. The “branches” of a water tree actually appear to comprise a high density of water-filled voids of typical diameter 1–10 μm . Such branches are therefore similar to a “string of pearls,” but in practice, even branches of water trees are not usually discernible. They are simply diffuse regions of water-filled voids. If dried

up, re-immersion in water reopens the voids. Boiling stabilizes the structure but probably also produces extra small voids. There is limited evidence that a percolation network does interconnect the voids, but the size scale of the interconnecting features is at around 10 nm. Electrolyte material accompanies the water into the voids and the ability of cationic dyes, such as rhodamine B, to stain the trees permanently indicates that some oxidation must have taken place. Chemical modification has also been shown using IR and FTIR spectroscopy and by fluorescence techniques.

Water trees grow much more slowly than electrical trees. Typically, they may not be observed at all for several years, even if the prevailing conditions for their growth are in place. They will then grow fast initially and then very slowly. Indeed, in the case of bow-tie water trees, there is much evidence that they stop growing completely after a given length



Underground Cable Systems. Figure 7

Contour plot of temperatures increment ($^{\circ}\text{C}$) above base-case temperature for selected base temperatures and moisture content. The calculations have been made for three difference cable voltages

(dependent on prevailing conditions) and that they might not precipitate breakdown. Vented water trees may cross the insulation completely without breakdown occurring, but they do greatly weaken the insulation. Generally, an electrical tree or a breakdown path may grow back through an electrical tree.

It is important to recognize that water trees occur in all extruded insulations (EPR, WTRXLPE, XLPE) and have been found in failure locations retrieved from service. It is often suggested that the water tree-retarding insulations (EPR and WTRXLPE) do not grow water trees. Unfortunately, this is not correct, though it is more difficult to detect water trees in these insulations; however, this is due to the lower initiation and growth rates (EPR and WTRXLPE) and the opaque nature of the insulation (EPR). Nevertheless, longer endurances and lower failure rates are seen for EPR and WTRXLPE for comparable designs and ages, with respect to cable XLPE analogues.

There are many proposed mechanisms of water treeing and these have been critically reviewed in Reference [12]. Essentially, it is likely that solvated ions are injected at partially oxidized sites. These catalyze further oxidation by maintaining the ion concentration. A sequence of metal-ion-catalyzed reactions is proposed in which bonds break and cause microvoids to develop. Alternating electromechanical stresses open up pathways for solvated ions; these initiate new microvoids. Many tree-retardant polymers contain “ion catchers” to prevent the metal ion catalysis, and these have been found to successfully delay the onset and growth rate of water trees.

The tree inception time, i.e., the time between the conditions being right for water tree growth and the first observation of water trees, is highly dependent upon the electrical stress. Typically, the inception time is inversely proportional to a high power (≈ 4 – 10) of electric field. For this reason, low-voltage cables, which tend to run at lower electric fields, may not have a water barrier to prevent water ingress and hence electrical treeing. In such cases, with fields typically < 4 kV/mm (see Table 3), the probability of failure through electrical treeing is low and a water barrier would make little difference. HV and EHV polymer-insulated cables generally use water barriers, and these become mandatory above 66 kV. Furthermore, the conductor is often water blocked (a water-swellaable compound or an extruded

mastic) to prevent the transport of water along the conductor. The water may enter the conductor either after a cable breakdown or during installation or through an incorrectly installed accessory.

Dry Aging: Thermoelectric Aging

The requirement for extra high voltage (EHV) underground power cables is increasing [1, 2, 12, 14, 22, 23]. There is commercial pressure to push the mean electric field in the insulation of such cables toward 16 kV/mm, and the most common insulation used is cross-linked polyethylene (XLPE). Long-term experience of XLPE, however, is limited to moderately stressed cables with mean fields of 5–7 kV/mm. Furthermore, the introduction of cross-linking processes has permitted the continuous operating temperature of polymeric cables (XLPE and EPR) to be increased to 90°C, equaling that of oil-filled (LPOF and HPOF) paper and polypropylene paper laminate (PPLP) cables. The use of XLPE as the insulation for transmission cables has grown steadily since the early 1990s. Many extruded power cables have been operating for 20 years and are approaching the end of their 30-year design life. If robust methodologies could be found for improving or/and evaluating the reliability of AC power cables, it may be possible to continue to use them without compromising the reliability of the system. Such methodologies require considerable improvements in the understanding of any aging or degradation mechanisms of cable insulation. They would enable XLPE cables to be more competitive at EHV levels.

Future Directions

When looking at the history of underground power cable systems, it is possible to discern a number of rather constant trends:

- Longer length systems get installed as service experience increases.
- Utilities are looking for technologies with easier and less expensive installation and maintenance.
- Manufacturing and material technologies permit higher operating stresses leading to reduced wall thicknesses.
- Reduced wall thicknesses:

- Lower the total installed system costs
- Improve the thermal capacity
- Place greater strains on the accessory and cable installation practices
- End users are looking for longer lives, though the end of life criteria remain undefined.
- Higher reliabilities for systems.
- Public pressure on the location of transmission lines underground.
- Increasing use of underground distribution cables in urban and suburban areas.

It seems clear that all of these trends will continue in the foreseeable future.

Considering the growth of energy requirement (Fig. 8), the increasing environmental awareness and the ever-growing level of information transfer, it can be seen that there will be a need to widen our future considerations to include:

- Much lower levels of electrical losses within the transmission and distribution systems – required for increased efficiency
- Higher attention to electric and magnetic field issues – required for increased acceptance of cable systems
- Even higher levels of urban undergrounding to improve the visual environment – required for increased acceptance of electric energy [25, 26]

- Integration of data and energy transmission – required for increased control and optimal use of corridors
- Improved levels of power delivery reliability as there is increasing reliance on more electrically powered systems – required for increased customer satisfaction

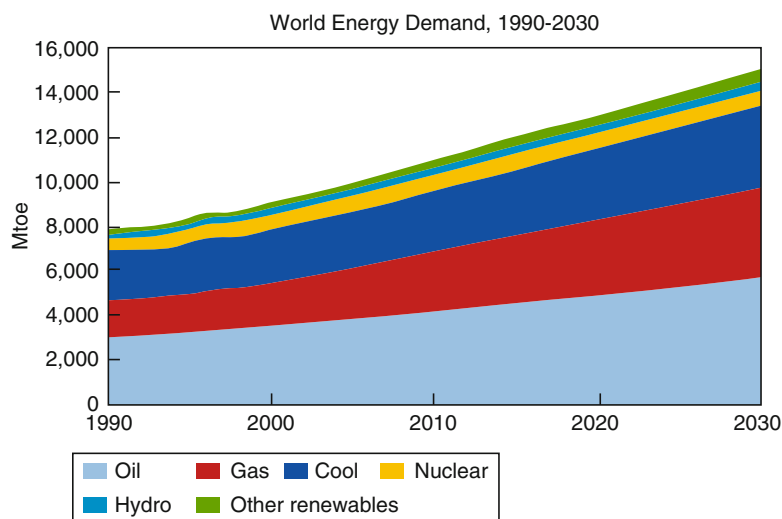
There can be no doubt that the use and importance of cables will increase.

The areas where most activity will be seen are:

- Understanding life expectancy of cable systems
- Improving cable performance with respect to aging
- Recyclable/recoverable cable designs
- Increased use of long length links
- Diagnostic trends for cable systems
- Impact of smart grid initiatives
- High temperature superconductivity (HTS)
- Gas-insulated lines (GIL)

Understanding Life Expectancy of Cable Systems

Utilities the world over continue to strive to increase the useful life of their underground or subsea cable system assets. However, this activity is taking place in an environment where there is an absence of any actuarial life expectancy or a good understanding of the factors that determine the end of useful life of a cable



Underground Cable Systems. Figure 8

Estimated world energy demand including split of primary energy sources [24]

system. This does not mean that some of the causes of failure are not understood, in fact an extensive amount of work has been done on early failures in the 15–30 year range (e.g., water treeing, effects of contaminants). Thus, the life estimate for cable systems of 40 years is difficult to justify in any rigorous way.

A good example of the issues is the case of MV PILC cables installed in the USA [11]. These cables make up 15% of the total US installed capacity; however, in certain critical locales, this can rise to 80%. The median age of the oldest of these cables is 80 year with lower and upper quartiles at 70 and 90 years. Yet these cables are not failing at a rate that is discernibly higher than the average population of PILC cables: median age of 44 years. Thus, it is not possible to classify them as reaching the end of life because they are old or when they have not reached the right-hand portion of the “bathtub curve.” It is equally unhelpful to assert that these cables have an infinite life as it is known that the paper and oil components will inexorably degrade.

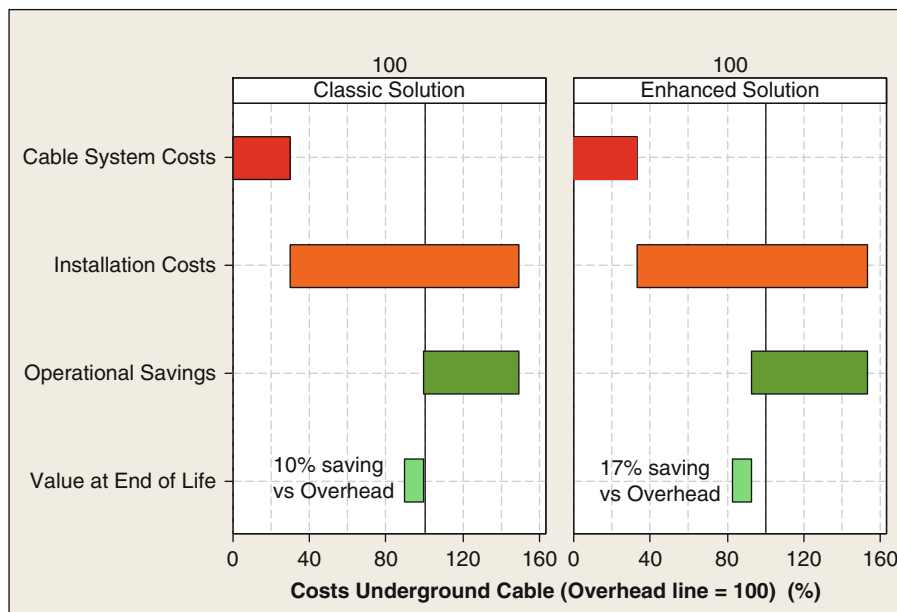
These concerns become important when considering the economics/reliability and sustainability of any

technology: The impact of a system that might last for 60 years is considerably less than today’s arbitrary estimate of 30–40 years.

Improving Cable Performance with Respect to Aging

If one builds on the life expectancy concepts discussed previously, then it becomes possible to ascribe a value to technologies that might increase that longevity. Figure 9 shows the financial benefits of improved reliability within MV system. The experience over the recent past has shown that the most efficient way to measure and assure cable reliability is to require long-term wet aging tests (CENELEC – 2 years and ICEA – 1 year) with success levels that comfortably exceed the specified minimums.

The total cost perspective of two cable installations is shown in Fig. 9. The first shows the classic reference scenario, where the installed cost of cables is higher than the overhead analogue but the total lifetime cost is lower (92%). The second shows the scenario where a longer-lived cable has a longer life (30 years rather than



Underground Cable Systems. Figure 9

Total cost perspective of two cable installations when compared to an Overhead Line. Initial costs are significantly offset by operational (maintenance, losses, and unreliability – all of these are assumed to be constant over a 30-year period) costs. Two cases are shown: 25 year life is the base case; the other is where a 10% higher cable cost brings higher quality and a 20% increase in life length (30 years)

Underground Cable Systems. Table 10 110kV Cable versus overhead cost (excluding permitting cost and time). Studies from The Technical University of Graz [26]

Type	Loading conditions	Installed costs (Euros/km)	Operating costs (Euros/km)	Total lifetime costs (Euros/km)	Cost ratio cables to overhead lines
Overhead lines	Low Load	149,000	15,000	164,000	1.18
Underground cable		181,000	12,000	193,000	
Overhead lines	High Load	149,000	165,000	314,000	0.76
Underground cable		181,000	58,000	239,000	

25 years in the reference) and lower operational costs, due to the better reliability throughout the longer life. Clearly the total cost is lower (82%) for the longer-lived cable. This approach is even more favorable than shown here for the higher-quality, longer-lived, cable; this is because the reference scenario would require that the cable and installation costs be incurred again in years 25–30.

A similar story is seen at HV: In Austria, a very detailed study of total lifetime costs at 110 kV (Table 10) shows that cables have lower costs compared to overhead lines when the normal high loading of lines is considered. The ratios favor overhead lines at low loading but the gap is likely to narrow when the cost of obtaining wayleaves and negative customer/regulator pressure is included [26].

Recyclable/Recoverable Cable Designs

Today there is little activity associated with the recovery of abandoned cable systems as it is believed that there is little economic value associated with them. However, this may well change in the future as:

- There is likely to be public and environmental pressure to rehabilitate the land.
- The cables represent considerable natural resources.
- There will be a pressure to reuse the land previously used for cables and more importantly their associated substations/transformers/switches.

Evidence of these trends can be seen in the recovery of submarine fluid-filled cables (such as those that previously lay in Long Island Sound); here the regulation

authorities required their removal as a prerequisite to the installation of new links. This reuses the structures can be seen with the replacement of gas compression cables in Germany and other places with extruded cables; but the reuse of the existing pipes/containment structures.

One of the limitations on recycling is the mixed nature of the nonmetallic components, i.e., the insulations and jackets [27]. This is because it is relatively simple to separate metals and metals from paper/plastic. However the plastic separation is not straightforward. Thus, to enable this separation in the future, the key element is the use of a whole polyolefin concept: XLPE and a HDPE jacket. With this approach the material mass is low and there is no expensive (\$50/T at 2003 costs) separation step at the end of useful life. The metal within the cable may be reused, and the energy content of the polymer will be liberated and provide a value extremely close to the prevailing cost of oil, which is likely to be significant in 25–30 years' time [27].

Increased Use of Long Length Links

Submarine Cable Systems

Submarine power cables is the term given to cables that carry power underwater. They may be major transmission systems running under the sea, river crossings, or links to islands. Most submarine cable systems, of short to moderate length, are AC, whereas very long lengths employ DC (see later). The reason for the preference for AC is the simplicity of integration with existing infrastructures. The AC embodiments suffer from

reactive losses, due to the natural capacitance and inductive properties of wire; hence, there are limitations on length. However, considerable accommodations can be made with solid state control devices and reactive compensation. DC transmission does not suffer reactive losses; the losses in the DC transmission line are the resistive losses. However, the losses in the AC/DC converters need to be included. Furthermore, the converter stations are a considerable capital and maintenance cost. However, from a pure power delivery standpoint, a DC cable system will carry between 1.3 and 1.6 times the power of an equivalently sized AC analogue.

Submarine cables are considerably longer than their terrestrial analogues: hundreds of kilometers versus tens of kilometers. They are laid in very long lengths from a variety of barges or ships. Thus, the manufacture and reliability requirements are much more challenging than for land cables. The importance of these considerations is obvious when the actions for repair after a failure in service are considered. The parties involved need to mobilize considerable marine resources, locate suitable repair materiel, obtain necessary environmental permits, physically attend to the site, retrieve the cable from the ocean floor, effect a repair, and confirm the integrity of the repair. If these elements are not challenging enough, the user and manufacturer need to determine some means to reestablish their confidence in the future problem-free operation of the link. Thus the design, specification, testing, manufacture and installation phases of a submarine project are many orders of magnitude more onerous than a similar terrestrial solution.

Nevertheless, in the future, it is clear that power grids will install larger amounts of submarine cable systems; the drivers will be:

- Increased network interconnection and stability
- Power trading
- Elimination of costly and inefficient local generation (islands)
- Integration of hydro and wind energy
- Reduced public reaction with respect to terrestrial power links

Efficient DC Power Transmission

Direct current (DC) power transmission has been shown to be highly efficient for highly controlled

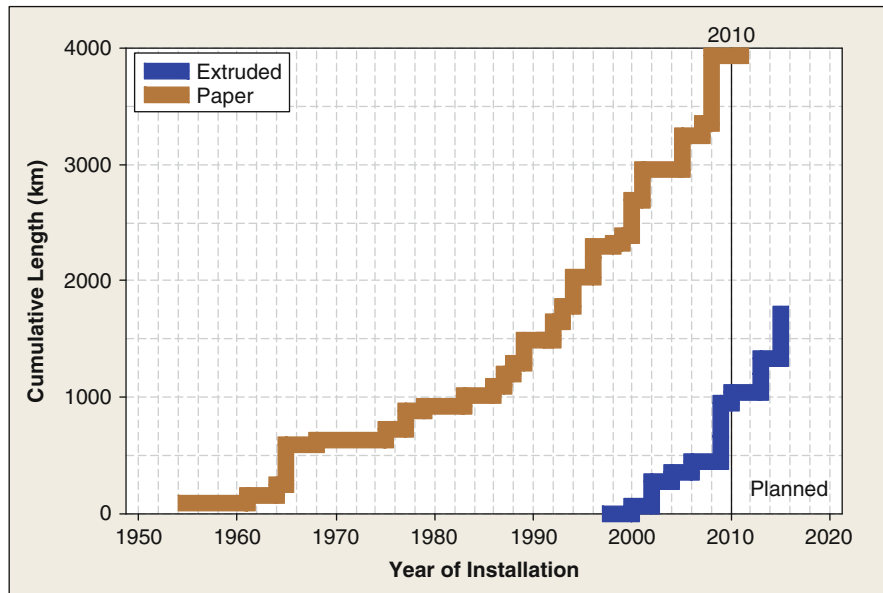
long-distance delivery (experience in Nordpool). Long-distance AC transmission is possible (Horns Rev and Isle of Man); however, it requires complex system control and reactive compensation of the cable capacitance. Thus, today the only technology that is capable of delivering long-distance power utilizing underground cables is HVDC. Inspection of the Gotland, Cross Sound, Murraylink, Troll and Estlink projects shows us that voltage source converter (VSC) technology and cables manufactured with cross-linked polyethylene designed for the rigors of DC are already proven and in commercial use. It is interesting to note how many appliances used today operate with DC power supplies (PC, TVs, etc.) – perhaps Thomas Edison had it right all along!!

In the early days of DC, the increased cost of the converters limited its use to EHV grid interconnections (UK/France, Baltic Cable, etc.). However, recent innovations (VSC) in converter technology, cross-linked polyethylene cables, and the flexibility of system design have seen a very rapid growth of DC at high voltage (50–150 kV).

There have been interesting laboratory studies and qualifications of extruded cables using filled insulations at EHV; however, all of the world's commercial installations and the experience shown in Fig. 10 has been achieved with specially designed unfilled cross-linked materials. It is clear that unfilled cross-linked technology will be used when extruded cables are integrated into EHV DC systems.

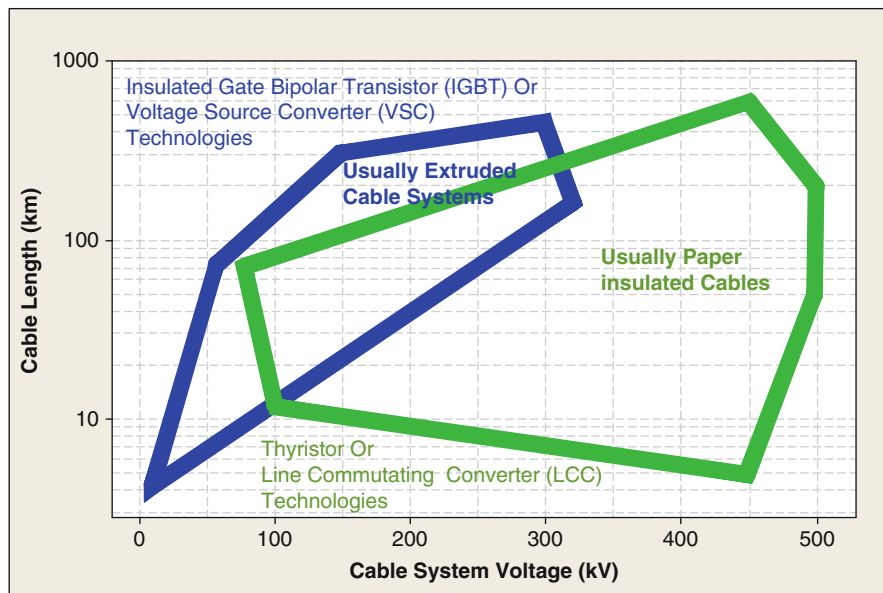
The increased use of DC is being supported within the international committees. Specific recommendations have been prepared and published by CIGRE WG 21.01 within Technical Brochure 219. This document provides a very solid base for the extension to extruded HVDC cables. DC cable design presents the engineer with a coupled thermal and electrical problem with the stresses experienced by the cable depending upon the temperature and stress (polarity) inversions. There can be no doubt that HVDC cable design presents as many challenges as EHV ac systems. These include:

- Stress inversions
- Electrical stresses higher than those seen at EHV AC
- Ultralong lengths
- Testing and approvals
- Accessory technology
- System integration



Underground Cable Systems. Figure 10

Global installed DC cable capacity segregated by type of cable insulation



Underground Cable Systems. Figure 11

Global installed DC cable capacity segregated by converter technology

However, as with EHV AC, careful attention to the insulation system makes it possible to succeed with some very impressive projects (Murraylink, Troll, and Cross Sound).

A complimentary aspect is the increasing usage of extruded cables using cross-linked DC polyethylene and voltage source converters (VSC) (Figs. 10 and 11). These developments have significantly increased the

speed of implementation, lowered the total cost, and increased the types of projects for DC systems. It is interesting to reflect in Fig. 10 that paper systems required 30 years to achieve 1,000 km of cumulative installed capacity whereas extruded cables within 13 years. It is clear that the trend toward extruded cable systems with VSC technology will increase both in the submarine and terrestrial environments.

Diagnostic Trends for Cable Systems

The use of diagnostics on cable systems is growing and will clearly be an important part of network management for the future. New cables and accessories will probably have features integrated into them that will provide significant amounts of data. Older systems do not benefit from these developments, and thus, a diagnosis will need to be made without optimal sensors and a baseline condition. Nevertheless, considerable advances have been made. However, the challenge for cable and system engineers will come in transforming these data into useful information. Although the new installations will bring exciting opportunities, the major challenge will be associated with the “dumb” cable systems of today and yesterday.

The major areas of activity for diagnostics are:

- Real-time control of cable system rating – This is normally achieved through the use of fiber-optic temperature systems which are coupled to sophisticated thermal models of the cable system.
- Cable system commissioning/acceptance tests – Cable system components are separately tested in the factory; however, there is overwhelming consensus that most of the cable system issues are associated to incidents that occur during installation. Thus, there is a growing trend for cable system acceptance tests which permit these defects to be detected prior to acceptance for service, thereby enabling repair in a timely and cost-effective manner.
- Cable system diagnostic tests [11] – These are tests where the “health” of the cable system is determined, together with essentially a probabilistic assessment of future performance. These results enable a proactive level of asset optimization within the operator.

Real-time control of cable system and cable system commissioning tests are well established, and mature approaches and implementation is proceeding through international standardization bodies such as CIGRE and IEC. In many cases, the challenge is not technical but is the availability of the appropriate knowledge within the user community on the appropriate implementation.

Cable System Diagnostic Tests Almost all electric power utilities distribute a portion of the electric energy they sell via underground cable systems. Collectively, these systems form a vast and valuable infrastructure. Estimates indicate that underground cables represent 15–20% of installed distribution system capacity. Utilities have a long history of using underground system with some of these cable systems installed as early as the 1920s. Very large quantities of cable circuits were installed in the 1970s and 1980s due to the introduction of economical, polymer-based insulation compounds and the decreasing acceptance of overhead distribution lines. Today, the size of that infrastructure continues to increase rapidly as the majority of newly installed electric distribution lines are placed underground.

Cable systems are designed to have a long life with high reliability. However, the useful life is not infinite. These systems age and ultimately reach the end of their reliable service lives. Estimates set the design life of underground cable systems installed in the range of 30–40 years. Today, a large portion of this cable system infrastructure is reaching the end of its design life, and there is evidence that some of this infrastructure is reaching the end of its reliable service life. This is a result of natural aging phenomena as well as the fact that the immature technology used in some early cable systems is decidedly inferior compared to technologies used today. Increasing failure rates on these older systems are now adversely impacting system reliability, and it is readily apparent that action is necessary to manage the consequences of this trend.

Complete replacement of old or failing cable systems is not an option. Many billions of dollars and new manufacturing facilities would be required. Electric utilities and cable/cable accessory manufacturers are simply not in a position to make this kind of investment.

However, complete replacement of these systems may not be required because cable systems do not age uniformly. Cable researchers have determined that many cable system failures are caused by isolated cable lengths or isolated defects within a specific circuit segment. Thus, the key to managing this process is to find these “bad actors” and to proactively replace them before their repeated failures degrade overall system reliability. Various cable system diagnostic testing technologies were developed to detect cable system deterioration. The results of diagnostic tests are used to identify potential failures within cable systems and then again, after repair, to verify that the repair work performed did indeed resolve the problem(s) detected.

Appropriate maintenance and repair practices enable system aging to be controlled and help manage end-of-life replacements. Diagnostics to determine the health of the cable system are critical to this management program.

A number of cable diagnostic techniques are now offered by a variety of service providers and equipment vendors. However, no one service has definitively demonstrated an ability to reliably assess the condition of the wide variety of cable systems currently in service. Implementing cable system diagnostics in an effective way involves the management of a number of different issues. This includes the type of system (network, loop, or radial), the load characteristics (residential, commercial, high density, government, health care, etc.), the system dielectric (XLPE, EPR, paper, mixed), and system construction (direct buried or conduit). The basic cable diagnostic testing technologies used to assess cable circuit conditions are listed below.

- Time domain reflectometry (TDR)
- Partial discharge (PD) at operating, elevated 60 Hz, elevated Very Low Frequencies (VLF) or damped AC (DAC) Voltages [11, 28]
- Tan δ /dielectric spectroscopy at 60 Hz, VLF or variable frequencies [28–30]
- Recovery voltage
- DC leakage current
- Polarization and depolarization current
- Simple withstand tests at elevated VLF, 60 Hz AC, or DC Voltages [11, 31]
- Acoustic PD techniques
- Monitored withstand tests at elevated VLF, 60 Hz AC, or DC voltages with simultaneous monitoring of PD, tan δ , or Leakage Current [11, 31]
- Combined diagnostic tests at 60 Hz AC, very low frequencies (VLF), or damped AC (DAC) voltages using PD and tan δ

There is no doubt that cable system diagnostic testing can be used to improve system reliability. However, to be effective, the technology should be appropriate to the circuit to be tested. Setting accurate and reasonable expectations is also a critical part of the process.

In general, the work performed in the CDFI [11] led to the following observations:

- Diagnostic tests can work. They often show many useful things about the condition of a cable circuit, but not everything desired.
- Diagnostics do not work in all situations. There are times when the circuit is too complex for the diagnostic technology to accurately detect the true condition of the circuit.
- Diagnostics are generally unable to determine definitively the longevity of the circuit under test. Cable diagnostics are much like medical diagnostics. They can often tell when something is wrong (degraded), but it is virtually impossible to predict the degree to which a detected defect will impact the life of the system tested.
- Field data analysis indicates that most diagnostic technologies examined do a good job of accurately establishing that a cable circuit is “good.” They are not as good at establishing which circuits are “bad.” In most cases, there are far more good cable segments than bad segments. However, it is virtually impossible to know which “bad” circuits will actually fail. Therefore, utilities must act on all replacement and repair recommendations to achieve improved reliability.
- The performance of a diagnostic program depends on:
 - Where diagnosis is used?
 - When diagnosis is used?
 - Which diagnosis to use?
 - What is done afterward?
- A quantitative analysis of diagnostic field test data is very complex. The data comes in many

different formats and the level of detail is extremely variable. However, an in-depth analysis of the data clearly highlights the benefits of diagnostic testing.

- Diagnostic data require skilled interpretation to establish how to act. In almost all cases, the tests generate data requiring detailed study before a decision can be made on whether to repair or replace the tested cable circuit.
- No one diagnostic is likely to provide sufficient information to accurately establish the condition of a cable circuit.

Impact of Smart Grid Initiatives

Much has been written about Smart Grid Initiatives. Smart grid is defined rather differently in Europe and USA; however, one commonality is that the goal is to insert intelligent and interactive devices in the existing grid infrastructure to fulfill a number of goals, such as:

- Better facilitate the connection and operation of generators of all sizes and technologies
- Allow consumers to play a part in optimizing the operation of the system
- Significantly reduce the environmental impact of the whole electricity supply system
- Maintain or even improve the existing high levels of system reliability, quality, and security of supply
- Dynamic optimization of grid operations and resources, with full cyber-security
- Deployment and integration of distributed resources and generation, including renewable resources
- Development and incorporation of demand response, demand-side resources, and energy-efficiency resources
- Deployment of “smart” technologies (real-time, automated, interactive technologies that optimize the physical operation of appliances and consumer devices) for metering, communications concerning grid operations and status, and distribution automation
- Deployment and integration of advanced electricity storage and peak-shaving technologies, including plug-in electric and hybrid electric vehicles, and thermal-storage air conditioning

Thus, it is clear that the existing grid will be operated in a different manner to the one that it is accustomed and that the demands placed upon it will be increased. A natural consequence will be that a number of hitherto unseen failure or degradation modes will become prevalent. It is already possible to postulate what a number of them might be:

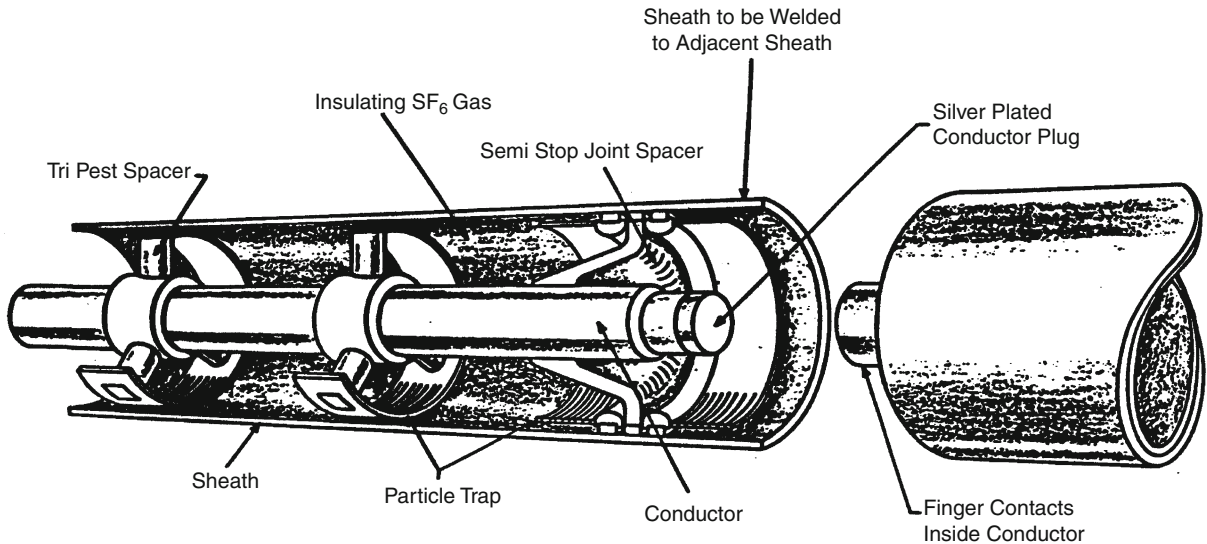
- Overheating of nonoptimally designed/installed components, which have not presented themselves due to the lower level of grid loading [32]
- Accelerated degradation due to transients/harmonics superposed on the grid by the smart devices themselves, the most obvious are third, and higher, harmonics
- A heightened sensitivity of consumers to power quality issues
- More aggressive load profiles due to changing demand and supply side protocols, as a consequence of reducing the traditional “low load recovery periods”

High-Temperature Superconductivity (HTS)

Equally there are technologies of today that will find some niche uses, but are unlikely to come into immediate and widespread use. The first is high-temperature superconductivity (HTS) that may well find application where people wish to transmit large amounts of power over relatively short distances. However, the issues associated with long-distance cryogenics, termination temperature differences, and complicated start-up procedures need to be addressed before this technology becomes attractive to utilities.

Gas-Insulated Lines (GIL)

Gas-insulated Lines offer the hope of large current ratings, low capacitances, and reduced dielectric losses for longer transmission distances. Yet, in most practical cases, the short lengths, and the associated difficulties of fabrication, plus the large environmental issues around large amounts of SF₆ gas, even in mixtures, will limit its use. By comparison, it would seem that GIL has much wider application than superconducting cables. A typical gas-insulated line design is shown in Fig. 12.



Underground Cable Systems. Figure 12
Typical design of a gas-insulated line (GIL)

Bibliography

- Moore GF (1997) *Electrical cables handbook*. Blackwell Science, San Diego. ISBN 0 632 04075 0
- Densley J (1995) Ageing and diagnostics in extruded insulations for power cables. In: IEEE 5th international conference on conduction and breakdown in solid dielectrics, IEEE publication 95CH3476-9, pp 1–15
- Fothergill JC, Hampton RN (2004) Polymer insulated cable. In: Haddad A, Warne D (eds) *Advances in high voltage engineering*. IEE, London, pp 477–507, Chapter 10. ISBN 0 85296 158 8
- Abernethy RB (1996) *The new Weibull handbook*, 2nd edn. Robert B. Abernethy, North Palm Beach. ISBN 0965306208
- Hampton RN (2008) Some of the considerations for materials operating under High Voltage Direct Current Stresses. *IEEE Electrical Insulation Magazine* 24(1):5–13
- Leijon M (1999) Novel concept in high voltage generation: Powerformer™. IEE Conference Publication, v5, n467, 1999. In: *Proceedings of the 1999, 11th International symposium on "High Voltage Engineering"*, 23–27 Aug 1999, London, UK, p 5.379.S5-5.382.P5, IEE, ISSN: 0537-9989 IECPB4 ISBN: 0-85296-719-5
- Jenkins N (2002) *Power Engineering Journal* 16(3):90–91
- Karlstrand J, Sunnegardh P, Zenger W, Ghafurian R, Boggia R (2002) Water-cooled 345 kV solid-dielectric cable system, Cigré paper 21–111
- Halvarsson P, Karlstrand J, Larsson D, Larsson M, Reinholdsdóttir K, Sigurðsson E (2002) A novel approach to long buried AC transmission system, Cigré paper 21–201
- Bartnikas R, Eichhorn RM (eds) (1983) *Engineering dielectrics*, vol IIA, *Electrical properties of solid insulation*, Philadelphia, ASTM-STP 783
- Diagnostic Testing of Underground Cable Systems (Cable Diagnostic Focused Initiative) DOE Award No. DE-FC02-04CH11237, December 2010
- Dissado LA, Fothergill JC (1992) Electrical degradation and breakdown in polymers. In: Peter Peregrinuns for the IEE, London
- Brincourt T, Regaudie V (1999) Evaluation of different diagnostic methods for the French underground MV network. In: JICABLE'99, Versailles, p 451
- Peschke E, Schroth R, Olshausen R (1995) Extension of XLPE cables to 500 kV based on progress in technology. In: JICABLE'95, Montreal, pp 6–10
- Griffiths CL, Betteridge S, Hampton RN (1998) Thermoelectric ageing of cable grade XLPE in dry conditions. *Conduction and breakdown in solid dielectrics ICSD '98. Proceedings of the 1998 IEEE 6th international conference*, pp 279–282, doi:10.1109/ICSD.1998.709279
- Coppard RW, Bowman J, Dissado LA, Roland SM, Rakowski RT (1990) The effect of aluminium inclusions on the dielectric breakdown strength of polyethylene. *J Appl Phys* 23:1554–1561
- Antoniski JR, Nilsson U, Gubanski SM (1997) The effect of metal inclusions on the breakdown strength of XLPE cables. In: CEIDP 1997, Minneapolis, pp 283–286
- Hagen ST, Ildstad E (1993) Reduction of AC breakdown strength due to particle inclusions in XLPE cable insulation. In: *Conference on power cables and accessories 10 kV–500 kV*, London, pp 165–168

19. Blodgett RB (1962) Influence of absorbed water and temperature on tan delta and dielectric constant of oil-impregnated paper insulation. *Trans AIEE* 81:93–100
20. Hartlein RA, Hampton RN, Perkel J (2008) Some considerations on the selection of optimum location, timing, and technique, for diagnostic tests. In: IEEE Power Engineering Society (PES) general meeting panel session, Pittsburgh
21. Steennis EF (1989) Water treeing: the behaviour of water trees in extruded cable insulation. ISBN 90-353-1022-5, Kema, pp 132–133
22. Gregory B, Griffiths CL, Hampton RN and Mainwaring SP (2000) A probabilistic design method for polymeric structures to operate at high electrical stress. In: 8th international conference on dielectric materials, measurements and application, IEE Conference Publication no 473, pp 419–424
23. Ogawa K, Kosugi T, Kato N, Kawawata Y (1990) The worlds first use of 500 kV XLPE insulated aluminium sheathed power cables at the Shimogo and Imaichi power stations. *IEEE Trans Power Delivery* 5(1):26–32
24. International Energy Agency. www.iea.org
25. Gorham and Partners (1995) Undergrounding – A success story already in Europe – Potentially worldwide. Report to ICF Congress 1995
26. Campbell A (2003) Undergrounding HV and EHV – window of opportunity opens. In: ICF Congress 2003, St Petersburg, pp 169–193
27. Hampton RN (2005) Electrical insulation systems for energy technology and clean environment. In: Nordic insulation symposium (NORDIS), Trondheim, June 2005
28. Hampton RN, Olearczyk M, Perkel J, Weisenfeld N (2010) Cable fleet management. *IEEE Power and Engineering*, Nov 2010
29. Perkel J, Hernández JC, Hampton RN, Drapeau JF, Densley (2011) Challenges associated with the interpretation of dielectric loss data from power cable system measurements. In: International conference on insulated power cables JICABLE11, Versailles France, June 2011, C.4.5
30. Drapeau J-F, Cantin B, Hampton N, Perkel J (2011) Dielectric loss characterization of hydro-quebec medium voltage (MV) submarine cables – phase 1. In: International conference on insulated power cables JICABLE11, Versailles France, June 2011, Paper A.7.6
31. Fletcher CL, Perkel J, Hampton RN, Hernandez JC, Hesse J, Pearman MG, Wall CT, Zenger W (2011) First practical utility implementation of monitored withstand diagnostics in the USA. In: International conference on insulated power cables JICABLE11, Versailles France, June 2011, Paper A.10.2
32. Parker T, Hampton RN (2011) Impact of filled-strand conductor on connector temperature for medium voltage joints. In: International conference on insulated power cables JICABLE11, Versailles France, June 2011, Paper C.1.5

Unscented Kalman Filter in Intelligent Vehicles

MOUSTAPHA DOUMIATI¹, DANIEL LECHNER²

¹B2i Automotive Engineering Company, Massy, France

²Department of Accident Mechanism Analysis, IFSTTAR-MA Laboratory, Salon de Provence, France

Article Outline

Glossary

Definition of the Subject

Introduction

Estimation Process Description

Four-Wheel Vehicle Model

Tire-Road Interface

Observers Design

Experimental Results

Conclusion and Future Perspectives

Acknowledgments

Bibliography

Glossary

Lateral tire forces They are responsible to hold on the vehicle during a turn.

Longitudinal tire forces They are responsible to accelerate/brake the vehicle.

Observer or estimator It models a real system in order to provide an estimate of its internal state, given measurements of the input and output of the real system.

Sideslip angle It is the angle between the velocity heading and the true heading of the vehicle.

Tire forces The developed forces (longitudinal and lateral) are function of tire properties (material, tread pattern, tread depth, profile, etc.), the normal load on the tire, and the velocities experienced by the tire.

Vehicle control systems They provide commands and instructions to control the movements of the vehicle in order to maintain stability and enhance passengers security and comfort.

Vehicle dynamics It includes analytical and experimental technology used to study and understand

the dynamical responses of a vehicle in various in-motion situations.

Vertical (normal) tire forces They are responsible to support the weight of the vehicle.

Definition of the Subject

The principal concerns in driving safety with standard vehicles are understanding and preventing risky situations. A close examination of accident data reveals that losing the vehicle control is the main reason for most car accidents. To help the driver to prevent such accidents, vehicle control systems may be used. For their optimal operation, these control systems require certain input data concerning vehicle dynamic parameters and vehicle–road interaction. Unfortunately, some fundamental parameters like the tire-road forces and the sideslip angle are difficult to measure in a car, for both technical and economic reasons. To face this problem, this study presents a dynamic modeling and observation method to estimate these variables. The ability to accurately estimate lateral tire forces and sideslip angle is a critical determinant in the performances of many vehicle control systems. To address nonlinearities and unmodeled vehicle dynamics, an observer derived from unscented Kalman filtering technique is proposed. The estimation process method is based on the dynamic response of a vehicle instrumented with easily available and potentially integrable sensors. Performances are tested using an experimental car in real driving situations. Experimental results show the potential of the proposed estimation method.

Introduction

Vehicle dynamics and stability have been of considerable interest to automotive engineers, automobile manufacturers, government, public safety groups, and general public for a number of years. The obvious dilemma is that people naturally desire to drive faster and faster on the roads and highways, yet they expect their vehicles to be stable and safe during all normal and emergency maneuvers. For the most part, people pay little attention to the limited handling potential of their vehicles until some unusual behavior is observed that often results in fatality. According to statistics, worldwide, an estimated 1.2 million people are killed

in road crashes each year and as many as 50 million are injured [1]. Preventing car accidents requires to know what determines vehicle dynamics during motion [2].

Today, automotive electronic technologies are developing for safe and comfortable traveling of drivers and passengers. Nowadays, there are a lot of vehicle control system such as the Anti-lock Braking System (ABS) that prevents wheel lock during braking [3], and the Electronic Stability Control (ESC) that enhances lateral vehicle stability [4, 5]. These control systems installation rate is increasing all around the world. Vehicle control algorithms have made great strides toward improving the handling and safety of vehicles. For example, experts estimate that ESC prevents 27% of loss-of-control accidents by intervening when emergency situations are detected [6]. While nowadays vehicle control algorithms are undoubtedly a lifesaving technology, they are limited by the available vehicle state information.

Vehicle control systems currently available on production cars rely on available inexpensive measurements such as longitudinal velocity, accelerations, and the vehicle yaw rate. Sideslip rate can be evaluated using the yaw rate, lateral acceleration, and vehicle velocity [7]. Calculating the sideslip angle is possible from the sideslip rate integration. However, it is prone to uncertainty and errors from sensor bias. Besides, these control systems use unsophisticated, inaccurate tire models to evaluate lateral tire dynamics. In fact, measuring tire forces and sideslip angles is very difficult for technical and economic reasons. Therefore, these important data must be observed or estimated. If control systems were in possession of the complete set of lateral tire characteristics, namely lateral forces, sideslip angle, and the tire-road friction coefficient, they could greatly enhance vehicle handling and increase passenger safety.

As the motion of a vehicle is governed by the forces generated between the tires and the road, knowledge of the tire forces is crucial when predicting vehicle motion. For example, a vehicle can turn because of the applied lateral tire forces. In fact, what happens is that when the front wheels of a vehicle are steered, a slip angle is created, which gives rise to a lateral force. This lateral force turns or yaws the vehicle. Under normal driving situations (low slip angle), a vehicle responds predictably to the driver's inputs. As the vehicle

approaches the handling limits, for example, during an evasive emergency maneuver, or when a vehicle undergoes high accelerations, high slip angle occurs, and the vehicle's dynamic becomes highly nonlinear and its response becomes less predictable and potentially very dangerous.

Accurate data about tire forces lead to a better evaluation of the vehicle possible trajectories, to a better vehicle control and rollover prevention. Moreover, it enables the development of a diagnostic tool for evaluating the potential risks of accidents related to poor adherence or dangerous maneuvers.

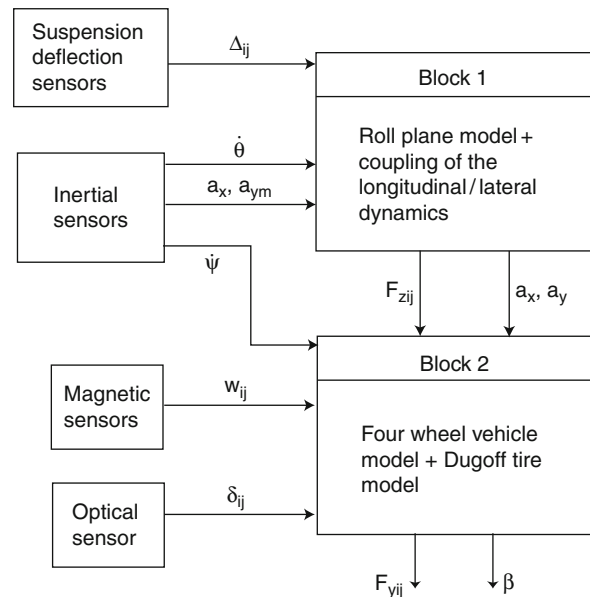
In the literature, many studies have looked at the vehicle dynamic states estimation. Several ones have been conducted regarding the estimation of tire/road forces [8–20]. For example, in [8], a study of a 14-degrees-of-freedom (DOF) vehicle model is proposed where the dynamics of the roll center are used to calculate vertical tire forces. In [9], the authors propose an estimation method in order to estimate vertical and lateral forces per axle. The authors in [10–13] estimate the vehicle vertical forces and other dynamic states for a four-wheel vehicle model (FWVM) comprising four DOF. Consequently, lateral tire forces at each tire are calculated based on the estimated states and using a quasi-static tire model. In [14], Ray estimates the vehicle dynamic states and lateral tire forces per axle for a nine-DOF vehicle model. The author uses measurements of the applied torques as inputs to his model. We note that the torque is difficult to get in practice; it requires expensive sensors. More recently, authors in [15, 16] propose observers to estimate lateral forces per axle without using torque measures. In [17], the authors propose an estimation process based on a three-DOF vehicle model, as a lateral tire force estimator. In [15–17], lateral forces are modeled with a derivative equal to random noise. The authors in [17] remark that such modeling leads to a noticeable inaccuracy when estimating individual lateral tire forces, but not in axle lateral forces. This phenomenon is due to the non-representation of the lateral load transfer when modeling. Studies in [21–24] focus on the tire–road friction estimation.

The main goal of this study is to present an estimation method that uses simple vehicle-road models and a certain number of valid measurements in order to estimate in real time and in accurate way the sideslip

angle and the lateral force at each individual tire/road contact point. This study presents two significant particularities (Fig. 1):

1. First, the estimation process does not use the measurements of wheel torques which are very expensive.
2. Second, the estimation process uses accurate normal tire forces, in contrast to many existing approaches that assume constant vertical forces. This approach is more realistic since during cornering, accelerating, and braking, the load distribution varies significantly in a car, thus cornering stiffness and lateral forces evaluation are directly affected.

The developed estimation process is model-based and built using Kalman filter technique. The Kalman filter is known as the most commonly used real-time estimator for linear and nonlinear systems. In order to show the effectiveness of the estimation method, some real-time validation tests were carried out on an instrumented vehicle in realistic driving situations.



Unscented Kalman Filter in Intelligent Vehicles. Figure 1
Process estimation diagram

Estimation Process Description

The estimation process is shown in its entirety by the block diagram in Fig. 3, where a_x and a_{ym} are, respectively, the longitudinal and lateral accelerations; $\dot{\psi}$ is the yaw rate, $\dot{\theta}$ is the roll rate, Δ_{ij} (the index i represents the front (1) or the rear (2) and the index j represents the left (1) or the right (2)) is the suspension deflection; w_{ij} is the wheel velocity; F_{zij} and F_{yij} are, respectively, the normal and lateral tire-road forces; β is the sideslip angle at the center of gravity (cog). The whole estimation process consists of two blocks, and its role is to estimate sideslip angle at the cog, normal, and lateral forces at each tire/road contact point, and consequently evaluate the used lateral friction coefficient. The following measurements are needed:

- Yaw and roll rates measured by gyrometers
- Longitudinal and lateral accelerations measured by accelerometers
- Suspension deflections using suspension deflections sensors
- Steering angle measured by an optical sensor
- Rotational velocity for each wheel given by magnetic sensors

The first block aims to provide the vehicle's mass, lateral load transfer, normal tire forces, and the corrected lateral acceleration a_y (by canceling the gravitational acceleration component that distorts the accelerometer signal a_{ym}). It contains observers based on vehicle's roll dynamics and model that couples longitudinal and lateral accelerations. Authors have looked at the first block in some previous studies [10, 11]. This work focuses only on the second block, whose main role is to estimate individual lateral tire force and sideslip angle. The second block makes use of the estimations provided by the first block. In fact, as will be shown in the sections, the impact of including accurate normal forces in the calculation of lateral forces is fundamental.

One specificity of this estimation process is the use of blocks in series. By using cascaded observers, the observability problems entailed by an inappropriate use of the complete modeling equations are avoided, enabling the estimation process to be carried out in a simple and practical way. In the following, the model-based observer of the second block will be

explained in details. Therefore, the vehicle-road model and the estimation method will be illustrated, respectively.

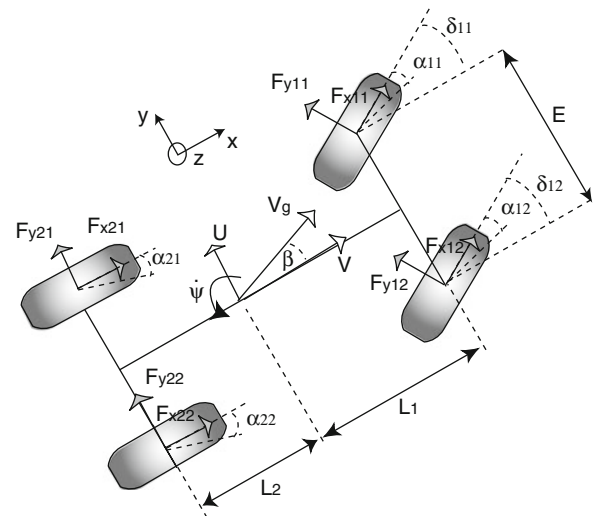
Four-Wheel Vehicle Model

The Four-Wheel Vehicle model (FWVM) is chosen for this study because it is simple and corresponds sufficiently to our objectives. The FWVM is widely used to describe transversal vehicle dynamic behavior [12–14, 17, 18].

Figure 2 shows a simple diagram of the FWVM model in the longitudinal and lateral planes.

In order to simplify the lateral and longitudinal dynamics, rolling resistance is neglected. Additionally, the front and rear track widths (E) are assumed to be equal. L_1 and L_2 represent the distance from the vehicle's center of gravity to the front and rear axles, respectively. The sideslip at the vehicle center of gravity (β) is the difference between the velocity heading (V_g) and the true heading of the vehicle (ψ). The yaw rate ($\dot{\psi}$) is the angular velocity of the vehicle about the center of gravity. The forward and lateral velocities are, respectively, V and U . The longitudinal and lateral forces ($F_{x,y,i,j}$) are shown for front and rear tires of the vehicle.

Longitudinal forces should be taken into account to enable accurate lateral forces estimation during vehicle



Unscented Kalman Filter in Intelligent Vehicles. Figure 2 Four-wheel vehicle model

braking or acceleration. While considering their effect is certainly important, its inclusion makes solving the lateral estimation problem considerably more complex. Thus, it may be desirable to solve the lateral estimation problem in the absence of longitudinal forces first and include them in later studies. This can be done by focusing on solving the estimation problem when the vehicle is driven at constant speeds [20].

This study extended the hypothesis of moving in a constant speed and addresses the case of a front-wheel drive, where rear longitudinal forces are neglected relative to the front longitudinal forces. Longitudinal front axle forces are considered by assuming that:

$$F_{x1} = F_{x11} + F_{x12} \quad (1)$$

The longitudinal force evolution is modeled with a random walk model, where its derivative is equal to random noise ($\dot{F}_{x1} = 0$). This is due to the lack of knowledge on the longitudinal slip and the effective radius of the tire [25].

The lateral dynamics of the vehicle can be obtained by summing the forces and moments about the vehicle's center of gravity. Consequently, the simplified FWVM is formulated as the following dynamic relationships [19]:

$$\begin{cases} \dot{V}_g = \frac{1}{m} \begin{bmatrix} F_{x1} \cos(\beta - \delta) + F_{y11} \sin(\beta - \delta) \\ F_{y12} \sin(\beta - \delta) + (F_{y21} + F_{y22}) \sin \beta \end{bmatrix} \\ \ddot{\psi} = \frac{1}{I_z} \begin{bmatrix} L_1 [F_{y11} \cos \delta + F_{y12} \cos \delta + F_{x1} \sin \delta] \\ -L_2 [F_{y21} + F_{y22}] \\ + \frac{E}{2} [F_{y11} \sin \delta - F_{y12} \sin \delta] \end{bmatrix} \\ \dot{\beta} = \frac{1}{mV_g} \begin{bmatrix} -F_{x1} \sin(\beta - \delta) + F_{y11} \cos(\beta - \delta) \\ F_{y12} \cos(\beta - \delta) + (F_{y21} + F_{y22}) \cos \beta \end{bmatrix} - \dot{\psi} \\ a_y = \frac{1}{m} [F_{y11} \cos \delta + F_{y12} \cos \delta + (F_{y21} + F_{y22}) + F_{x1} \sin \delta] \\ a_x = \frac{1}{m} [-F_{y11} \sin \delta - F_{y12} \sin \delta + F_{x1} \cos \delta] \end{cases} \quad (2)$$

where m is the vehicle mass, and I_z is the yaw moment of inertia.

The tire slip angle (α_{ij}), as shown in Fig. 2, is the difference between the tire's longitudinal axis and the tire's velocity vector. The tire velocity vector can be obtained from the vehicle's velocity (at the cog) and

the yaw rate. Assuming that rear steering angles are approximately null, the direction or heading of the rear tires is the same as that of the vehicle. The heading of the front tires includes the steering angle (δ). The front steering angles are assumed to be equal ($\delta_{11} = \delta_{12} = \delta$). The forward velocity V , steering angle δ , yaw rate $\dot{\psi}$, and the vehicle body slip angle β are then used to calculate the tire slip angles α_{ij} , where:

$$\begin{cases} \alpha_{11} = \delta - \arctan \left[\frac{V\beta + L_1\dot{\psi}}{V - E\dot{\psi}/2} \right] \\ \alpha_{12} = \delta - \arctan \left[\frac{V\beta + L_1\dot{\psi}}{V + E\dot{\psi}/2} \right] \\ \alpha_{21} = -\arctan \left[\frac{V\beta - L_2\dot{\psi}}{V - E\dot{\psi}/2} \right] \\ \alpha_{22} = -\arctan \left[\frac{V\beta - L_2\dot{\psi}}{V + E\dot{\psi}/2} \right] \end{cases} \quad (3)$$

Tire–Road Interface

As the motion of a vehicle is governed by the forces generated between the tires and the road, knowledge of the tire forces is crucial in order to predict the vehicle's motion. This section presents the tire–road interaction phenomenon, especially the lateral tire forces. Since the quality of the observer largely depends on the accuracy of the tire model, the underlying model must be precise. Taking real-time calculation requirement, the tire model should also be simple.

Dugoff Tire Model

Many different tire models, based on the physical nature of the tire and/or on empirical formulations deriving from experimental data, can be found in the literature. These models include the Burckhardt, Dugoff, and Pacejka models [19, 26, 27]. One of the most commonly used model is the Pacejka's "Magic Formula." It is an effective method for predicting real tire behavior. However, it requires a large number of tire-specific parameters that are usually unknown. Another commonly used model is the Dugoff tire model. It synthesizes all the tire property parameters into two constants C_x and C_y , referred to as the longitudinal and cornering stiffness of the tire. Dugoff's model is the one used in this study. Neglecting

longitudinal forces, the simplified nonlinear lateral tire forces are given by:

$$\overline{F_{yij}} = -C_{yij} \tan \alpha_{ij} f(\lambda) \quad (4)$$

where C_{yij} is the cornering stiffness, α_{ij} is the slip angle, and $f(\lambda)$ is given by:

$$f(\lambda) = \begin{cases} (2 - \lambda)\lambda, & \text{if } \lambda < 1 \\ 1, & \text{if } \lambda \geq 1 \end{cases} \quad (5)$$

$$\lambda = \frac{\mu F_{zij}}{2C_{yij} |\tan \alpha_{ij}|} \quad (6)$$

In the above formulation, μ is the friction coefficient and F_{zij} is the vertical load on the tire. This simplified tire model assumes pure slip conditions with negligible longitudinal slip, a uniform pressure distribution, a rigid tire carcass, and a constant friction coefficient for sliding rubber. The original Dugoff tire model has a constant stiffness in respect to weight transfer. It is worthy to note that, according to [28], load transfer affects the cornering stiffness C_{yij} . It can be represented by a second-order polynomial with respect to the normal force as shown in (7):

$$C_{yij}(F_z) = (aF_{zij} - bF_{zij}^2) \quad (7)$$

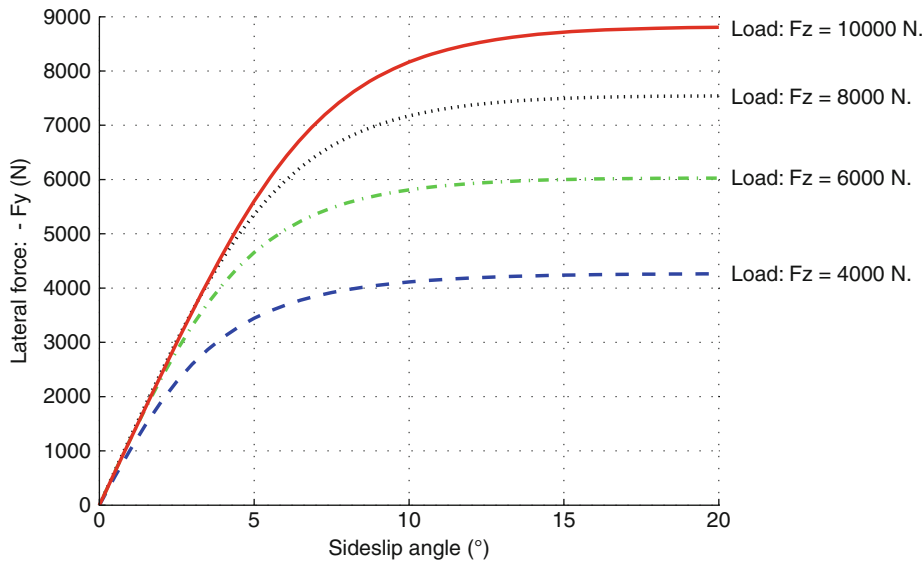
where a and b are, respectively, the first and second-order coefficient in the cornering stiffness polynomial. They are identified once using a set of experimental data treated offline, where the tires remain in their linear operation zone. Therefore, (7) is compared to the calculated cornering stiffness obtained from the ratio of the measured lateral force to the measured tire's slip angle. Hence, a and b are calculated, in such a way that the (7) fits the calculated cornering stiffness well.

This study proposes a modified Dugoff tire model, where the cornering stiffness varies with respect to load.

As shown in (4), vertical forces and the tire slip angles can be used to find the lateral force on each tire. Figure 3, based on the modified Dugoff's tire model, is a graph of the lateral force versus tire slip angle. It will be noted that as the load increases, the peak lateral force occurs at somehow higher slip angle.

It is clear that for small slip angles, the force profile can be defined by a linear region. When operating in this region, a vehicle responds predictably to the driver's inputs.

As the slip angle continues to grow, the tire begins to saturate and reaches a peak value; this area is commonly called the nonlinear region of the tire curve. It represents the tire limits, and it is rarely reached



Unscented Kalman Filter in Intelligent Vehicles. Figure 3
Generic tire curve: lateral force versus slip angle

under normal driving conditions. If the front tires saturate first, the vehicle is said to display understeer and may plow out of a bend. If the rear tires saturate first, the vehicle limit oversteers and may spin out. Because most drivers are not accustomed to operate in the nonlinear handling regime, both of these responses are potentially very dangerous. Noting that oversteer situation is much more difficult to be controlled than the understeer [29].

Observers Design

This section presents a description of the observer devoted to lateral tire forces and sideslip angle. The state-space formulation, the observability analysis, and the estimation method will be presented.

Stochastic State-Space Representation

The nonlinear stochastic state-space representation of the system described in previous section is given as:

$$\begin{cases} \dot{X}(t) = f(X(t), U(t)) + w(t) \\ Y(t) = h(X(t), U(t)) + u(t) \end{cases} \quad (8)$$

The input vector, U , comprises the steering angle and the normal forces considered estimated by the first block (see Fig. 3):

$$U = [\delta, F_{z11}, F_{z12}, F_{z21}, F_{z22}]^T = [u_1, u_2, u_3, u_4, u_5]^T \quad (9)$$

The measure vector, Y , comprises yaw rate, vehicle velocity (approximated by the mean of the rear-wheel velocities calculated from wheel-encoder data), longitudinal and lateral accelerations:

$$Y = [\dot{\psi}, V_g, a_x, a_y]^T = [y_1, y_2, y_3, y_4]^T \quad (10)$$

The state vector, X , comprises yaw rate, vehicle velocity, sideslip angle at the cog, lateral forces, and the sum of the front longitudinal tire forces:

$$\begin{aligned} X &= [\dot{\psi}, V_g, \beta, F_{y11}, F_{y12}, F_{y21}, F_{y22}, F_{x1}]^T \\ &= [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8]^T \end{aligned} \quad (11)$$

Authors would like to emphasize that the consideration of the lateral forces as states allows:

1. A better evaluation of the tire forces. In fact, whatever the complexity of the tire models, there are

several reasons why such models do not match the actual tire forces perfectly [30]. From these reasons, we can cite especially the changes in the tire's pressure and temperature and the changes in the road characteristics. Therefore, authors believe that according to the closed loop observer theory, the integration of the tire forces in the state vector may lead to better results than just using an open loop tire model.

2. A better understanding of the tire behaviors using the relaxation-length formulation, especially in transient maneuvers [31].
3. The forces reconstruction to be done robustly with respect to some parameter variations. In fact, it is well known that the Kalman filters have proven to be robust to parameter changes.

Taking these observations in mind, one can infer the contribution of this study with respect to others existing studies in the literature like [12, 13], which estimate dynamic variables of the vehicle, and then assess the tire forces using a properly adjusted tire model.

The process and measurements noise vectors, respectively, w and u , are assumed to be white, zero mean, and uncorrelated.

Consequently, the particular nonlinear function $f(\cdot)$ of the state equations is given by:

$$\begin{cases} f_1 = \frac{1}{I_z} \begin{bmatrix} L_1 [x_4 \cos u_1 + x_5 \cos u_1 + x_8 \sin u_1] \\ -L_2 [x_6 + x_7] \\ + \frac{E}{2} [x_4 \sin u_1 - x_5 \sin u_1] \end{bmatrix} \\ f_2 = \frac{1}{m} \begin{bmatrix} x_8 \cos(x_3 - u_1) + x_4 \sin(x_3 - u_1) \\ + x_5 \sin(x_3 - u_1) + (x_6 + x_7) \sin(x_3) \end{bmatrix} \\ f_3 = \frac{1}{mx_2} \begin{bmatrix} -x_8 \sin(x_3 - u_1) + x_4 \cos(x_3 - u_1) \\ + x_5 \cos(x_3 - u_1) + (x_6 + x_7) \cos x_3 \end{bmatrix} - x_1 \\ f_4 = \frac{x_2}{\sigma_1} (-x_4 + \overline{F_{y11}}(\alpha_{11}, u_2)) \\ f_5 = \frac{x_2}{\sigma_1} (-x_5 + \overline{F_{y12}}(\alpha_{12}, u_3)) \\ f_6 = \frac{x_2}{\sigma_2} (-x_6 + \overline{F_{y21}}(\alpha_{21}, u_4)) \\ f_7 = \frac{x_2}{\sigma_2} (-x_7 + \overline{F_{y22}}(\alpha_{22}, u_5)) \\ f_8 = 0 \end{cases} \quad (12)$$

The observation function $h(\cdot)$ is:

$$\begin{cases} h_1 = x_1 \\ h_2 = x_2 \\ h_3 = \frac{1}{m}[-x_4 \sin u_1 - x_5 \sin u_1 + x_8 \cos u_1] \\ h_4 = \frac{1}{m}[x_4 \cos u_1 + x_5 \cos u_1 + (x_6 + x_7) + x_6 \sin u_1] \end{cases} \quad (13)$$

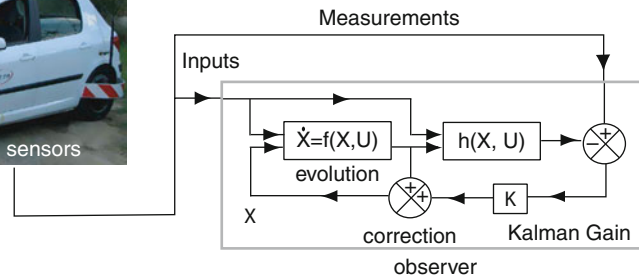
The state vector $X(t)$ will be estimated by applying the unscented Kalman filter technique.

Observability

Observability is a measure of how well the internal states of a system can be inferred from knowledge of its inputs and external outputs. This property is often presented as a rank condition on the observability matrix. Using the nonlinear state-space formulation of the system presented in [section “Stochastic State-Space Representation,”](#) the observability definition is local and uses the Lie derivative [32]. An observability analysis of this system was undertaken in [33]. It was shown that the system is observable except when:

- Steering angles are null.
- The vehicle is at rest ($V_g = 0$).

For these situations, we assume that lateral forces and sideslip angle are null, which approximately corresponds to the real cases.



Unscented Kalman Filter in Intelligent Vehicles. Figure 4
Process estimation diagram

Estimation Method: Kalman Filter Algorithms

The aim of an observer or a virtual sensor is to estimate a particular unmeasurable variable from available measurements and a system model in a closed loop observation scheme, as illustrated in [Fig. 4](#). Because of the vehicle system-model mismatches (unmodeled dynamics, parameter variations, etc.) and the presence of unknown and unmeasurable disturbances, the calculation obtained from vehicle's model would deviate from the actual values over time. In order to reduce the estimation error, at least some of the measured outputs are compared to the same variables estimated by the observer. The difference is fed back into the observer after being multiplied by a gain matrix K , and so we have a closed loop observer (see [Fig. 4](#)). All developed observers are implemented in a first-order Euler approximation discrete form. At each iteration, the state vector is first calculated according to the evolution equation and then corrected online with the measurement errors (innovation) and filter gain K in a recursive prediction–correction mechanism. In this study, the gain is calculated using the Kalman filter method which is a set of mathematical equations and is widely represented in [34–36].

The EKF (Extended Kalman Filter) is probably the most commonly used estimator for nonlinear systems. However, in this study, the UKF (Unscented Kalman Filter) [38–40] is chosen for the following fundamental reasons:

- The high nonlinearities of the vehicle-road model.
- The calculation complexity of the Jacobian matrices which causes implementation difficulties.

UKF Algorithm In this subsection, the principle of the UKF is summarized. Consider the general discrete nonlinear system:

$$\begin{cases} X_{k+1} = f(X_k, U_k) + w_k \\ Y_k = h(X_k) + v_k \end{cases} \quad (14)$$

where $X_k \in R^n$ is the state vector, $U_k \in R^r$ is the known input vector, $Y_k \in R^m$ is the output vector at time k , w_k and v_k are, respectively, the disturbance and sensor-noise vector, which are assumed to be Gaussian white noise with zero mean and uncorrelated.

The UKF can be formulated as follows [37–40]:

Initialization

$$\begin{cases} \bar{X}_0 = E[X_0] \\ P_0 = E[(X_0 - \bar{X}_0)(X_0 - \bar{X}_0)^T] \end{cases} \quad (15)$$

where $\bar{X}_0$ and P_0 are, respectively, the initial state and the initial covariance.

Sigma points calculation and time update

$$\begin{cases} \chi_{k-1} = [\bar{X}_{k-1}, \bar{X}_{k-1} \pm \sqrt{(n+\lambda)P_{k-1}}] \\ \chi_{k|k-1}^* = f(\chi_{k-1}, U_{k-1}) \\ \bar{x}_{k|k-1} = \sum_{i=0}^{2n} w_i^m \chi_{i,k|k-1}^* \\ P_{k|k-1} = \sum_{i=0}^{2n} w_i^c (\chi_{i,k|k-1}^* - \bar{x}_{k|k-1}) \times (\chi_{i,k|k-1}^* - \bar{x}_{k|k-1})^T + Q \\ \chi_{k|k-1} = [\bar{X}_{k-1}, \bar{X}_{k-1} \pm \sqrt{(n+\lambda)P_{k|k-1}}] \\ \gamma_{k|k-1} = h(\chi_{k|k-1}) \\ \bar{Y}_{k|k-1} = \sum_{i=0}^{2n} w_i^m \gamma_{i,k|k-1} \end{cases} \quad (16)$$

where

$$\begin{cases} w_0^m = \frac{\lambda}{n+\lambda} \\ w_0^c = \frac{\lambda}{n+\lambda} + (n - \alpha^2 + \beta) \\ w_i^m = w_i^c = \frac{1}{2(n+\lambda)} \quad i = 1, \dots, 2n \\ \lambda = n(\alpha^2 - 1) \end{cases} \quad (17)$$

Measurement update

$$\begin{cases} P_{\bar{Y}_k \bar{Y}_k} = \sum_{i=0}^{2n} w_i^c (\gamma_{i,k|k-1} - \bar{Y}_{k|k-1}) \times (\gamma_{i,k|k-1} - \bar{Y}_{k|k-1})^T + R \\ P_{\bar{X}_k \bar{Y}_k} = \sum_{i=0}^{2n} w_i^c (\chi_{i,k|k-1} - \bar{X}_{k|k-1}) \times (\gamma_{i,k|k-1} - \bar{Y}_{k|k-1})^T \\ K_k = P_{\bar{X}_k \bar{Y}_k} P_{\bar{Y}_k \bar{Y}_k}^{-1} \\ P_k = P_{k|k-1} - K_k P_{\bar{Y}_k \bar{Y}_k} K_k^T \\ \bar{X}_k = \bar{X}_{k|k-1} + K_k (Y_k - \bar{Y}_{k|k-1}) \end{cases} \quad (18)$$

where the variables are defined as follows: $\bar{X}_k$ and $\bar{Y}_{k|k-1}$ are the estimations, respectively, of the state and of the real measurement at each instant k , w_i is a set of scalar weights, and n is the state dimension; the parameter α determines the spread of the sigma points around $\bar{X}$ and is usually set to $1e-4 < \alpha < 1$. The constant β is used to incorporate part of the prior knowledge of the distribution of X , and for Gaussian distributions, $\beta = 2$ is optimal. Q and R are, respectively, the disturbance and sensor-noise covariance: R takes into account the uncertainty in the measured data and Q is tuned depending on the model quality. Remember that the computation of the Kalman gain is a subtle mix between process and observation noises. The less noise in the operation compared to the uncertainty in the model, the more the variables will be adapted to follow measurements. Since the lateral forces are modeled using a relaxation model based on reliable tire models, the uncertainty affected to them is not too high. However, the longitudinal force per front axle is not modeled at all; hence, it is represented by a high noise level. The other states (yaw rate, longitudinal and lateral vehicle velocity) are modeled using the vehicle's equations. Therefore, they are said to have an average noise. On the other hand, since the embedded sensors have good accuracy, the noises on the measurements are quite small. In order to reduce the complexity of the problem, both measurement covariance matrix, R , and the process covariance matrix, Q , are assumed to be constant and diagonal. The off-diagonal elements are set to 0. That means that both the measurement noises and the process noises are supposed uncorrelated.

Experimental Results

In this section, the experimental car used to evaluate the observer performances is presented. Moreover, the test conditions and the results of the previously developed observers are discussed and analyzed.

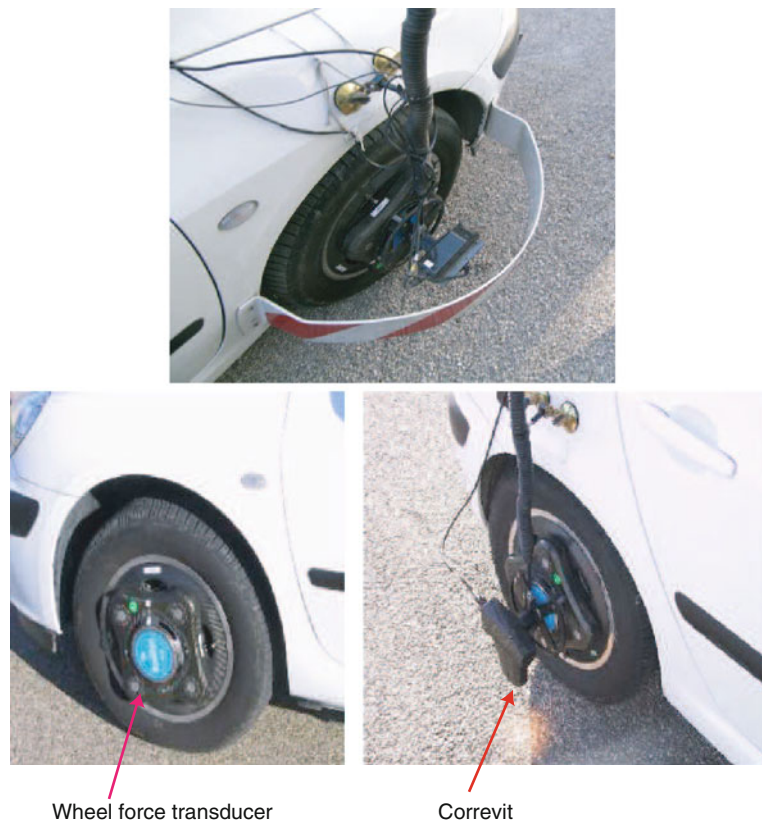
Experimental Car

The experimental vehicle shown in Fig. 4 is the INRETS-MA (Institut National de Recherche sur les Transports et leur Sécurité – Département Mécanismes d'Accidents) Laboratory's test vehicle [7]. It is a Peugeot 307 equipped with:

- Gyrometers and accelerometers that measure, respectively, the rotations (roll, pitch, and yaw rates) and accelerations (longitudinal, lateral, and vertical) of the car body.
- Suspension sensors that measure the distances between the wheels and the car body.

- Three Correvit noncontact optical sensors:
 1. One is located in chassis rear overhanging position, and it measures longitudinal and lateral vehicle speeds.
 2. The other two are installed on the front right and rear right tires, and they measure front- and rear-tires longitudinal/lateral velocities and sideslip angles.
- Dynamometric wheels fitted on all four tires, which are able to measure tire forces and wheel torques in and around all three dimensions.
- Steering-rack displacement sensor that is used to determine the steering angle.
- Magnetic sensors that measure rotational velocity for each wheel.

It is important to note that the Correvit and the wheel-force transducers (see Fig. 5) are very expensive sensors. They are used in this study as a reference for



Wheel force transducer

Correvit

Unscented Kalman Filter in Intelligent Vehicles. Figure 5
Wheel-force transducer and sideslip sensor installed at the tire level

validating the estimation process. The sampling frequency of the different sensors is 100 Hz.

Test Conditions

Test data from nominal as well as adverse driving conditions were used to assess the performance of the observer presented in [section “Observers Design,”](#) in realistic driving situations. We report a lane-change maneuver where the dynamic contributions play an important role. [Figure 6](#) presents the Peugeot’s trajectory (on a dry road), its speed, steering angle, and “g-g” acceleration diagram during the course of the test. The acceleration diagram, that determines the maneuvering area utilized by the driver/vehicle, shows that large lateral accelerations were obtained (absolute value up to 0.6 g). This means that the experimental vehicle was put in a critical driving situation.

The estimation process algorithm was written in C++ and has been integrated into the laboratory car as a DLL (Dynamic Link Library) that functions according to the software acquisition system.

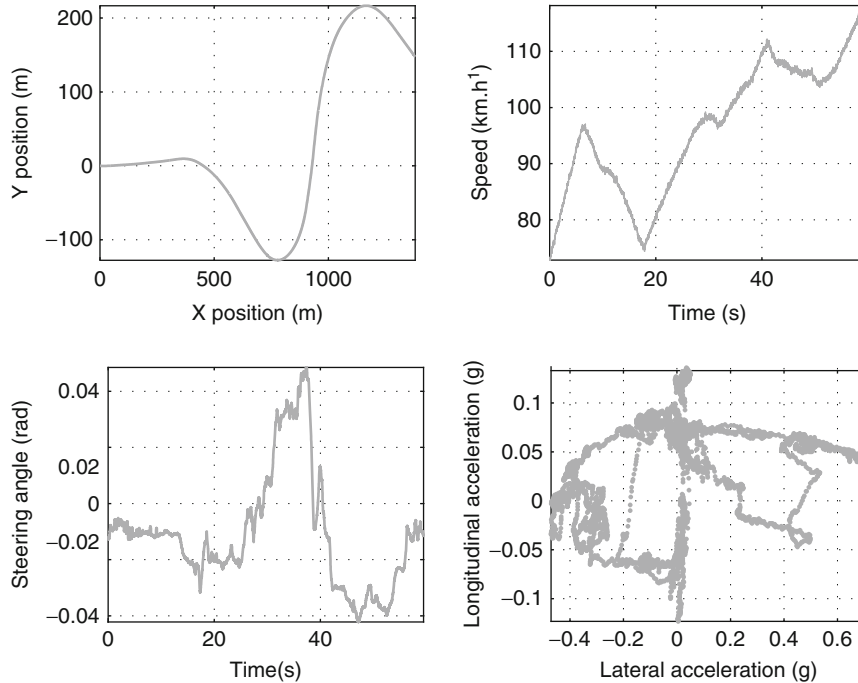
Validation of Observers

The observer results are presented in two forms: as tables of normalized errors, and as figures comparing the measurements and the estimations. The normalized error for an estimation z is defined as:

$$\epsilon_z = 100 \times \frac{\|z_{obs} - z_{measured}\|}{\max(\|z_{measured}\|)} \quad (19)$$

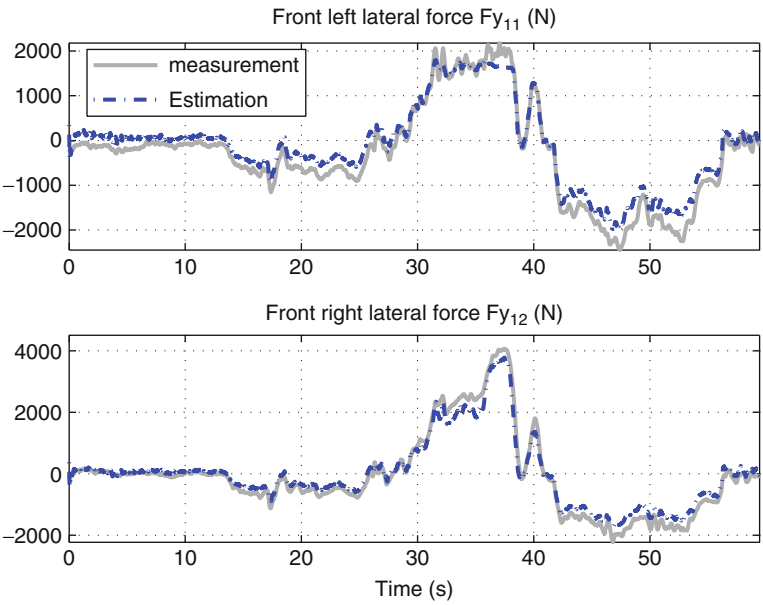
where z_{obs} is the variable calculated by the observer, $z_{measured}$ is the measured variable, and $\max(\|z_{measured}\|)$ is the absolute maximum value of the measured variable during the test maneuver.

[Figures 7](#) and [8](#) show lateral forces on the front and rear wheels. According to these plots, the observers are relatively good with respect to measurements. Some small differences during the trajectory are to be noted. These might be explained by neglected geometrical parameters, especially the camber angles, which also produce a lateral forces component [\[41\]](#). It is also shown that the lateral forces on the right-hand tires exceed those on the left-hand tires. This result is

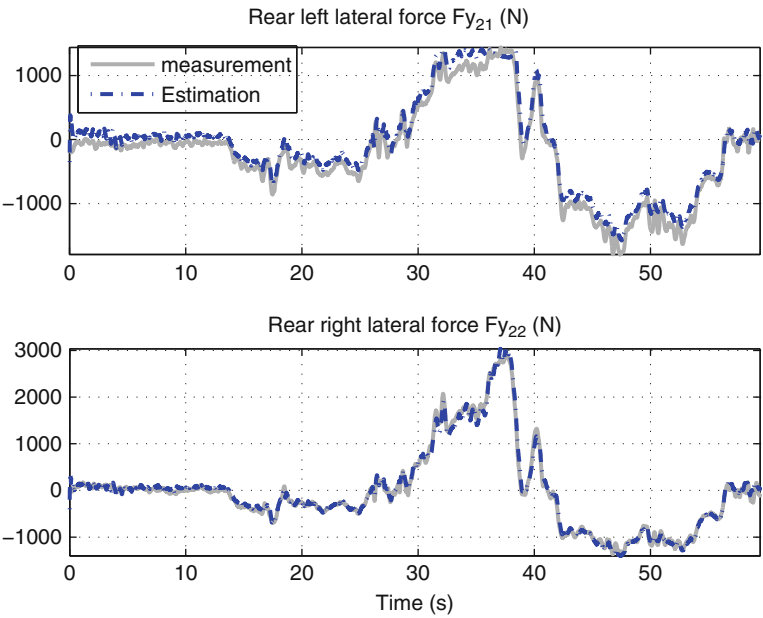


Unscented Kalman Filter in Intelligent Vehicles. **Figure 6**

Experimental test: vehicle trajectory, speed, steering angle, and acceleration diagrams for the lane-change test



Unscented Kalman Filter in Intelligent Vehicles. Figure 7
Estimation of front lateral tire forces

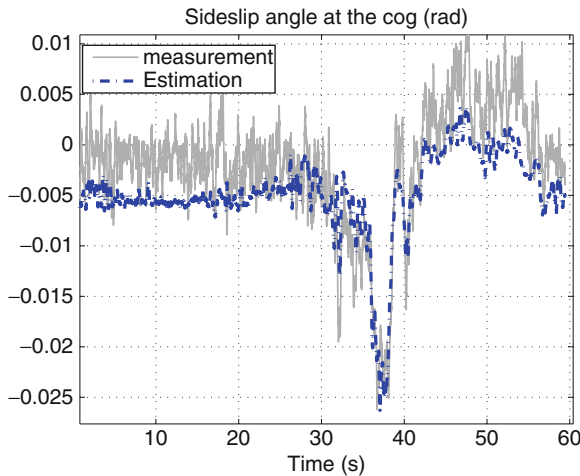


Unscented Kalman Filter in Intelligent Vehicles. Figure 8
Estimation of rear lateral tire forces

clearly a consequence of the load transfer produced during cornering from the left- to the right-hand side of the vehicle. In fact, lateral force increases as normal force increases. Figure 9 shows how sideslip

angle changes during the test. Reported results are relatively good.

Table 1 presents maximum absolute values, normalized mean errors, and normalized std (standard



Unscented Kalman Filter in Intelligent Vehicles. Figure 9
Estimation of the sideslip angle at the cog

Unscented Kalman Filter in Intelligent Vehicles.

Table 1 Maximum absolute values, normalized mean errors, and normalized Std

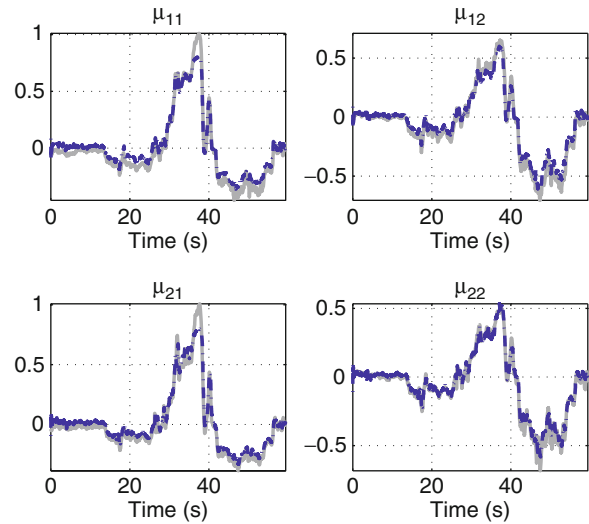
	Max	Mean%	Std%
F_{y11}	2180 (N)	8.23	4.80
F_{y12}	4070 (N)	3.70	3.74
F_{y21}	1441 (N)	7.51	3.52
F_{y22}	2889 (N)	1.91	1.77
β	0.027 (rad)	8.32	6.41

deviations) for lateral tire forces and sideslip angles. Despite the simplicity of the model, we can deduce that for this test, the performance of the observer is satisfactory, with normalized error globally less than 8%.

Given the vertical and lateral tire forces at each tire-road contact level, the estimation process is able to evaluate the used lateral friction coefficient μ . This is defined as the ratio of friction force to normal force, and is given by [29, 41]:

$$\mu_{ij} = \frac{F_{yij}}{F_{zij}} \quad (20)$$

The lateral friction coefficients in Figure 10 show that the estimated μ_{ij} are close to the measured values. A closer investigation reveals that the used lateral friction coefficients μ_{12} and μ_{22} corresponding to the



Unscented Kalman Filter in Intelligent Vehicles.

Figure 10

Used lateral friction coefficients developed by the tires

overloaded tires during cornering are lower than μ_{11} and μ_{21} . This phenomenon is due to the tire load sensitivity effect: the lateral friction coefficient is normally higher for the lighter loads, or conversely, falls off as the load increases [29, 41].

This test also demonstrates that μ_{11} and μ_{21} are high, especially for lateral accelerations up to 0.6, and that they attain the limit for the dry road friction coefficient. In fact, dry road surfaces show a high friction coefficient in the range 0.9–1.2 (implying that driving on these surfaces is safe), which means that for this test the limits of handling were reached.

Conclusion and Future Perspectives

This paper presents an interesting method for estimating lateral tire forces and sideslip angle, that is to say, two of the most important parameters affecting vehicle stability and the risk of leaving the road. Consequently, the developed observer could feed control systems with fundamental vehicle-dynamics data in order to enhance vehicle safety.

The proposed observer is derived from a simplified four-wheel vehicle model and is based on unscented Kalman filtering technique. Tire-road interaction is represented by the Dugoff model. A comparison with real experimental data demonstrates the potential of

the estimation process. It is shown that it may be possible to replace expensive correvit and dynamometric hub sensors by real-time software observers. This is one of the important results of this report. Another important result concerns the estimation of individual lateral forces acting on each tire. This can be seen as an advance with respect to the current vehicle-dynamics literature.

Future studies will improve the vehicle-road model in order to widen validity domains for the observer (take into account road irregularities and road bank angle), and make it adaptive with the road conditions (especially the road friction). Moreover, it will be of major importance to study the effect of coupling longitudinal/lateral dynamics on lateral tire behavior.

Acknowledgments

This work was done at Heudiasyc Laboratory UMR CNRS 6599, UTC University (Compiègne, France) in collaboration with Alessandro Victorino and Ali Charara.

Bibliography

Primary Literature

1. FARS (2007) Fatality analysis reporting system (FARS). Technical report, National Highway Traffic Safety Administration, National Center for Statistics and Analysis
2. Aparicio F, Paez P, Moreno F, Jimenez F, Lopez A (2005) Discussion of a new adaptive speed control system incorporating the geometric characteristics of the road. *Int J Veh Auton Syst* 3(1):47–64
3. Bosch R (2004) Bosch automotive handbook, 6th edn. SAE Society of Automotive Engineers, Bosch company, New York
4. ESC (2003) Electronic stability control fact sheet. Technical report, ESC Coalition and DEKRA Automotive research
5. Van Zanten A (2002) Evolution of electronic control systems for improving the vehicle dynamic behavior. In: *Proceedings of the 6th international symposium on advanced vehicle control (AVEC)*, Hiroshima, pp 7–15
6. Hsu Y, Gerdes JC (2005) Stabilization of a steerby-wire vehicle at the limits of handling using feedback linearization. In: *Proceedings of IMECE2005, AMSE international mechanical engineering congress and exposition*, Florida
7. Lechner D (2008) Embedded laboratory for vehicle dynamic measurements. In: *International symposium on advanced vehicle control*, Kobe
8. Shim T, Ghike G (2007) Understanding the limitations of different vehicle models for roll dynamics studies. *Veh Syst Dyn* 45:191–216
9. Doumiati M, Baffet G, Lechner D, Victorino A, Charara A (2008) Embedded estimation of the tire/road forces and validation in a laboratory vehicle. In: *International symposium on Advanced Vehicle Control*, Kobe
10. Doumiati M, Victorino A, Charara A, Lechner D, Baffet G (2008) An estimation process for vehicle wheelground contact normal forces. In: *IFAC World Congress*, Seoul Korea
11. Doumiati M, Victorino A, Charara A, Lechner D (2009) Lateral load transfer and normal forces estimation for vehicle safety experimental evaluation. *Veh Syst Dyn* 47(12):1511–1533
12. Wenzel TA, Burnham KJ, Blundell MV, Williams RA (2006) Estimation of the nonlinear suspension tyre cornering forces from experimental road test data. *Veh Syst Dyn* 44:153–171
13. Dakhlallah J, Glaser S, Mammari S, Seboudji Y (2008) Tire-road forces estimation using extended Kalman filter and sideslip angle evaluation. In: *American control conference*, Seattle
14. Ray LR (1997) Nonlinear tire force estimation and road friction identification: simulation and experiments. *Automatica* 33(10):1819–1833
15. Baffet G, Charara A, Dherbomez G (2007) An observer of tire road forces and friction for active-security vehicle systems. *IEEE/ASME Trans Mechatron* 12:651–661
16. Baffet G, Charara A, Lechner D, Thomas D (2008) Experimental evaluation of observers for tire-road forces, sideslip angle and wheel cornering stiffness. *Veh Syst Dyn* 45:191–216
17. Wilkin MA, Manning WJ, Crolla DA, Levesley MC (2006) Use of an extended Kalman filter as a robust tyre force estimator. *Veh Syst Dyn* 44:50–59
18. Venhovens PT, Naab K (1999) Vehicle dynamics estimation using Kalman filters. *Veh Syst Dyn* 32:171–184
19. Kiencke U, Daiss A (1997) Observation of lateral vehicle dynamics. *Control Eng Pract* 5(8):1145–1150
20. Hsu YHJ (2009) Estimation and control of lateral tire forces using steering torque. PhD, Department of mechanical engineering, Stanford University
21. Haffner L, Kozek M, Shi J, Jorgel HP (2008) Estimation of the maximum friction coefficient for a passenger vehicle using the instantaneous cornering stiffness. In: *American control conference*, Seattle
22. Gustafsson F (1997) Slip-based tire-road friction estimation. *Automatica* 33(6):1087–1099
23. Hahn JO, Rajamani R, Alexander L (2002) GPS-Based real time identification of tire-road friction coefficient. *IEEE Trans Control Syst Technol* 10(3):331–343
24. Uchanski MR (2001) Road friction estimation for automobiles using digital signal processing methods. PhD thesis, University of California, Berkeley
25. Koo SL, Tan HS (2007) Tire dynamic deflection and its impact on vehicle longitudinal dynamics and control. *IEEE/ASME Trans Mechatron* 12(6):623–631
26. Dugoff J, Fanches P, Segel L (1970) An analysis of tire properties and their influence on vehicle dynamic performance. *Society of Automotive Engineers Paper* 700377
27. Pacejka HB (2002) *Tyre and vehicle dynamics*. Elsevier, Amsterdam

28. Gillespie TD (1992) Fundamental of vehicle dynamics. Society of Automotive Engineers, Warrendale
29. Rajamani R (2005) Vehicle dynamics and control. Springer, New York
30. Lidner L (1993) Experience with the magic formula tyre model. In: Proceedings of the 1st international colloquium on tyre models for vehicle dynamic analysis, Amsterdam
31. Svendenius J (2007) Tire modeling and friction estimation. PhD thesis, Lund University, Sweden
32. Nijmeijer H, Van der Schaft AJ (1991) Nonlinear dynamical control systems. Springer, New York
33. Doumiati M, Victorino A, Charara A, Lechner D (2009) Unscented Kalman filter for real-time vehicle lateral tire forces and sideslip angle estimation. In: IEEE intelligent vehicles symposium, Xi'an, Shaanxi
34. Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME J Basic Eng 82(D):35–45
35. Welch G, Bishop G (2001) An introduction to the Kalman filter, Course 8. University of North Carolina, Department of computer science
36. Durrant-Whyte H (2001) Multi sensor data fusion. Technical report, Australian center for field robotics, University of Sydney
37. Mohinder GS, Angus PA (1993) Kalman filtering theory and practice. Prentice Hall, Englewood Cliffs
38. Julier SJ, Uhlman JK (1997) A new extension of the Kalman filter to nonlinear systems. In: International symposium aerospace/defense sensing, simulation and controls, Orlando
39. Julier SJ, Uhlman JK, Durrant-Whyte HF (1995) A new extension approach for filtering nonlinear systems. In: American control conference, Seattle
40. Wan EA, Merwe RVD (2000) The unscented Kalman filter for nonlinear estimation. In: Proceedings of the adaptive systems for signal processing, communications, and control symposium, Atlanta, pp 153–158
41. Milliken W, Milliken DL (1995) Race car vehicle dynamics. Society of Automotive Engineers, Warrendale

Books and Reviews

- Aga M, Okada A (2003) Analysis of vehicle stability control (VSC)'s effectiveness from accident data. In: Proceedings of the 18th international technical conference on the enhanced safety of vehicles, Nagoya
- Aguirre LA, Letellier C (2005) Observability of multivariate differential embeddings. J Phys A Math Gen 38:6311–6326
- Bakker E, Pacejka HB, Lidner L (1987) A new tire model with application in vehicle dynamic systems, SAE Paper 890087
- Besanon G (2007) Nonlinear observers and applications. Springer, Berlin
- Bevly DM, Gerdes JC, Wilson C, Zheng G (2000) The use of GPS based velocity measurements for improved vehicle state estimation: proceedings of the American Control Conference, Chicago. American Automatic Control Council, New York

- Brown T, Hac A, Martens J (2004) Detection of vehicle rollover. SAE World Congress, Detroit
- Doumiati M, Sename O, Martinez J, Poussot C, Dugard L (2010) Gain scheduled steering and braking control for improvement of vehicle handling and stability. In: IEEE conference on decision and control, Georgia
- Doumiati M, Sename O, Martinez J, Dugard L, Gaspar P, Szabo Z, Bokor J (2011) Vehicle yaw control via coordinated use of steering/braking systems. IFAC World Congress, Milano
- Heydinger G, Howe J (2000) Analysis of vehicle response data measured during severe maneuvers, SAE Paper 2000-01-1644
- Jazar RN (2008) Vehicle dynamics, theory and application. Springer, New York
- Smith ND (2004) Understanding parameters influencing tire modeling. Colorado State University, Formula SAE Platform

Uranium and Thorium Resources

J. STEPHEN HERRING

Idaho National Laboratory, Idaho Falls, ID, USA

Article Outline

Glossary
 Definition of the Subject
 Introduction
 Estimates of Uranium Reserves
 Global Estimates of Overall Uranium and Thorium Resources
 Mechanisms for the Concentration of Uranium
 Unconventional Resources
 New Technologies for Uranium Extraction
 Impact of Uranium Scarcity and Higher Extraction Costs
 Future Directions
 Bibliography

Glossary

Cross section Probability of neutron interaction with a nucleus, expressed in terms of area, in units of barns (b). One barn equals $1.0 \times 10^{-24} \text{ cm}^2$.

Enrichment The fraction of an isotope, usually fissile ^{235}U , in a mass of uranium. Enrichment is commonly quoted as the weight percent of the particular isotope. Natural uranium has an enrichment of

0.711 wt%, commercial reactor fuel is 3–5% enriched, and depleted uranium is 0.2–0.3% ^{235}U .

Enrichment tails (also depleted uranium) The uranium remaining after the enrichment of natural uranium into fuel, today about 0.3% ^{235}U , earlier 0.2–0.25% ^{235}U .

Fractionation Crystallization from a magma in which the initial crystals are prevented from equilibrating from the parent liquid, resulting in a series of residual liquids of more extreme composition than would have resulted from continuous reaction [1].

Highly enriched uranium (HEU) Uranium containing more than 20 wt% ^{235}U .

J_{th} Joule (i.e., Watt-second) thermal. One British thermal unit (BTU) equals 1,055 J_{th}.

Log-normal distribution Distribution of the form $f(x) = e^{-(\ln x)^2}$. In the present usage, the tonnage of an element available at concentration c , $T(c)$, is given by $T(c) = C_1 e^{-(\ln c_0 - \ln c)^2}$, where c_0 is the average crustal abundance and C_1 is a constant.

Low-enriched uranium (LEU) Uranium containing less than 20 wt% ^{235}U .

Mafic Composed chiefly of dark ferromagnesian minerals.

MOX Mixed oxide fuel, usually consisting of a ceramic mixture of uranium dioxide and plutonium dioxide.

MSWU Mega-separative work unit, a million separative work units. A separative work unit is the separative work that must be done to one kilogram of a mixture of isotopes to change its separation potential by one unit. The separation potential, a dimensionless function, is defined by $\phi(x_k) = (2x_k - 1) \ln \frac{x_k}{1-x_k}$, where x_k is the atomic fraction of the isotope, k . See Benedict 1981, p. 667 for a more complete definition [2].

Pegmatite An exceptionally coarse-grained igneous rock, with interlocking crystals, often found at the margins of batholiths.

Placer A mineral deposit at the surface formed by sedimentary concentration of heavy mineral particles from weathered debris.

Quad Quadrillion (i.e., 10^{15} , also written 1E15) British thermal units. One quad = 1.055×10^{18} J_{th}.

t Metric ton, also used in *Mt*, million metric tons, and *Tt*, trillion metric tons (teratons).

Unconformity A break or gap in the geologic record, such as an interruption in the normal sequence of deposition of sedimentary rocks, or a break between eroded metamorphic rocks and younger sedimentary strata [1].

Yellowcake A concentrate of uranium ore, containing 80–90% U_3O_8 . Yellowcake ranges from yellow to black, depending on impurities, processing temperature, and degree of hydration [3]. Although uranium prices are sometimes colloquially cited as “dollars per pound of yellowcake,” the actual prices are \$ per lb of U_3O_8 , where all of the uranium is assumed to be present in the yellowcake as that oxide.

Definition of the Subject

Uranium is a widely distributed element which is essential, at least in the near term, to the use of nuclear fission as a source of energy. Uranium is ubiquitous in the earth because of the wide variety of minerals in which it can occur, and because of the variety of geophysical and geochemical processes that have transported it since the primordial formation of the earth from the debris of supernovae. Uranium is approximately as common in the earth's crust as tin or beryllium, and is a minor constituent in most rocks and in seawater. The average crustal abundance of uranium is 2.76 weight parts per million (wppm), higher than the average concentrations of such economically important elements as molybdenum (1.5 wppm), iodine (0.5 wppm), mercury (0.08 wppm), silver (0.07 wppm), and gold (0.004 wppm).

Introduction

Beginning with the discovery of nuclear fission, uranium has been seen as a valuable but scarce resource. Uranium-235 (^{235}U) is the only naturally occurring isotope that can be made to fission with thermal neutrons. Consequently, the resources of uranium have been believed to inherently limit the sustainability of nuclear energy. There have been two periods of extensive exploration for uranium, in the 1950s and in the 1970s, both followed by long periods of severe contraction in the market and in exploration activity. With the peak of uranium prices to about \$350/kg in 2007, there

was an increased effort in exploration. However, that exploration quickly resulted in increased known reserves in several deposits and the return of prices to about \$140/kg. Today, exploration activity is at a moderate level for several reasons: (1) deposits found during earlier exploration periods have proven to be larger than initially estimated, (2) nuclear energy is growing, but not as rapidly as earlier forecast, (3) improved nuclear fuel management techniques and materials are allowing higher burnup and longer operating cycles, and (4) the conversion (“downblending”) of highly enriched uranium of military origin to civilian purposes has postponed the need for large amounts of newly mined natural uranium.

Concern that uranium would soon be exhausted was one of the driving forces in the development of fast breeder reactors, particularly in the 1960s and 70s. Fast breeder reactors convert the fertile isotope ^{238}U into the fissile isotope ^{239}Pu . These concerns also led to the development of thermal breeder reactors capable of converting thorium into the fissile, but not naturally occurring, isotope ^{233}U . Thorium, consisting almost entirely of the isotope ^{232}Th , is about four times more abundant than uranium and thus may represent a source of nuclear fuel in the distant future.

Estimates of Uranium Reserves

The importance of the overall uranium and thorium resource is demonstrated by the attention given to the estimates by such international agencies as the International Atomic Energy Agency (IAEA) and the Nuclear Energy Agency (NEA) of the Organisation for Economic Co-operation and Development (OECD). The estimates of these organizations are based on information provided by the member states and backed by research from others. The results are regularly compiled in the “Red Book” [4].

However, the estimates of uranium resources in the Red Book are based on the known and expected reserves that can be economically extracted using present or near-future technology. Because of the wide range of uranium concentrations in various minerals, the cost of extraction serves as the independent variable against which resources are estimated. Any reported amount of reserves/resources should be accompanied by the estimated cost of recovery of those reserves.

The Red Book [4] estimates the economically recoverable uranium reserves based on current and prospective mining projects (“Total Identified Resources Reasonably Assured and Inferred”) as of January 2009 are 5.4 million metric tons (Mt) of uranium of the best-proven category recoverable worldwide at a marginal cost of <\$130/kg of uranium metal. When the high-cost category (between \$130/kg and \$260/kg of U metal) is added, the total identified resources are estimated to be 6.3 Mt. Total undiscovered resources (prognosticated resources and speculative resources) as of January 2009 were estimated to be 10.4 Mt. The 2008 consumption of natural uranium by the 438 reactors worldwide was 59,065 t, and consequently, these total identified resources could be expected to last about 100 years at current rates and 250 years if speculative resources are included. These ratios of identified and speculative resources to consumption rate are longer than those for nearly all metals and fossil fuels, with the exception of coal.

Low market prices, the slow growth of nuclear power, and the downblending of highly enriched uranium (HEU) have in the past resulted in both very low levels of exploration and little effort in the development of advanced extraction technologies. Downblending agreements with the Russian Federation are due to expire in 2013. The discovery of additional total identified resources has shown a strong correlation with exploration expenditures, averaging 0.65 kg of additional resources per exploration dollar from 1987 to 2005 and 0.32 kg/exploration dollar from 2005 to 2009, following the uranium price increases after 2005 [5]. Those transient price increases led to an increase in the average annual expenditures for exploration from \$127 million/year for 1987–2005 to \$1.1 billion/year for 2005–2009.

It is important to note that uranium today is used overwhelmingly in the light water reactor fuel cycle, where only about 1.1–1.5% of the ultimate energy of the mined uranium is extracted via fissioning of ^{235}U and the small amounts of ^{239}Pu bred in situ. The rest of the uranium remains either in the used fuel or in the depleted uranium tails remaining after enrichment. Of the 1.8 Mt of uranium mined worldwide since 1945, the location of all but about 1,500 t is known. Only the location of that uranium dispersed either in nuclear explosions or as armor-piercing projectiles is not known. The inventory of used fuel and of depleted

uranium represents a very significant resource that could become fuel for fast breeder reactors.

While uranium is an essential input for the production of nuclear energy, the costs of natural uranium are a minor component of the overall cost. Today, at uranium prices of about \$140/kg U (\$53/lb of U_3O_8), the natural uranium required for the light water reactor (LWR) fuel cycle is responsible for only about 2.5–3% of generating costs. The fuel cost is 15–20% of the generating costs, but those costs include conversion of the uranium ore to UF_6 , enrichment of the natural uranium, production of the ceramic UO_2 fuel pellets, and fabrication of the fuel assemblies. A tenfold increase in the cost of natural uranium would not be welcome, but would not fundamentally change the economics of nuclear power. A tenfold increase in uranium prices would, at first estimate, be expected to increase the cost of nuclear electricity by 25–30%. However, a more detailed calculation, optimizing the ^{235}U content of the depleted uranium tails and adjusting fuel management for a higher priced resource, would result in an increase in the cost of electricity significantly less than 20%.

This entry discusses uranium resources in a global sense, beyond the official estimates of the IAEA and the OECD. As an introduction to that discussion, the origins of the earth's present inventory of uranium, the geophysical and geochemical processes that serve to concentrate uranium into economically viable deposits, and the technologies now being used for the extraction and concentration of uranium ores are described. Finally, the potential impacts of technologies now under development and the overall impact of the cost of uranium on the cost of energy from nuclear fission will be reviewed.

Thorium as a Nuclear Fuel

For the foreseeable future, uranium will probably continue to be the only source of nuclear energy. Nevertheless, for completeness, thorium resources should also be considered because of thorium's unique characteristics as a nuclear fuel. There are basically four reasons for considering thorium resources within the overall discussion of nuclear fuel resources. First, thorium is about 3.9 times more abundant than uranium, on a mass basis, as indicated both by samples of

the continental crust and by spectroscopy of supernova debris, from which planets are formed. Secondly, because similarities of the geochemistry and mineralogy of thorium and the lanthanides, thorium and the lanthanides (often called the "rare earth elements") are often found in the same ore bodies. Since the lanthanides are of increasing technical and strategic importance due to their widespread use in magnets and electronics, thorium is often treated as a waste since it has only a small market and since it is radioactive. Thus research on the efficient separation and purification of thorium could enhance both rare earth and thorium resources. Third, thorium can be directly substituted in the UO_2 crystal, making it a long-term supplement for uranium for in situ ^{233}U breeding. Thorium, as ThF_4 , can also be used in the molten salt reactor in combination with UF_4 , where the uranium would be a mixture of ^{233}U , ^{235}U , and ^{238}U . Finally, because thorium has only one oxide, ThO_2 , which has high thermal stability, it can serve as a very robust matrix for actinide transmutation, after which the ThO_2 would serve as the waste form for the transmutation targets.

Thorium averages 12 parts per million in the earth's crust, and is the 39th most abundant of the 78 crustal elements. Soil commonly contains an average of 6 wppm of thorium. When thorium is used as a nuclear fuel, much less plutonium and other minor transuranics (i.e., neptunium, americium, curium, berkelium, ...) are produced than are produced in uranium fuel cycles. The reduced production of transuranics occurs for two reasons. First, the fissile product of neutron absorption by ^{232}Th is ^{233}U , which is further down the actinide series from plutonium and the minor actinides, and secondly, the fission to capture ratio of ^{233}U is approximately nine, while that of ^{239}Pu is about three, thus resulting in lower production of the transuranics.

The generation of thorium deposits occurs in a fundamentally different manner from deposits of uranium. Uranium has some nine oxides and is dissolved or precipitated depending on the oxygen content and pH of the groundwater. Thus deposits are often formed where there is a decrease in the oxygen content of groundwater. Thorium, on the other hand, has only one oxide, ThO_2 , which is very refractory and insoluble. Thus thoria (along with many of the lanthanide oxides) is not dissolved in erosion by groundwater

and flowing rivers. The surviving grains, containing the thorium from the base rock, form into alluvial deposits of monazite sands.

Thorium occurs as the ores thorianite (ThO_2), thorianite (ThSiO_4) and mainly as monazite ((Ce, La, Nd, Th) PO_4). Thorium and its compounds have been produced primarily as a by-product of the recovery of titanium, zirconium, tin, and rare earths from monazite, which contains 6–8.5 wt% thorium oxide. Only a small portion of the thorium produced is consumed. Limited demand for thorium, relative to the demand for rare earths, has continued to create a worldwide oversupply of thorium compounds and mining residues. Most major rare-earth processors have switched feed materials to thorium-free intermediate compounds to avoid the handling of radioactive thorium. Excess thorium not designated for commercial use is either disposed of as a radioactive waste or stored for potential use as a nuclear fuel or other applications. Increased costs to comply with environmental regulations and potential legal liabilities and costs to purchase storage and waste disposal space were the principal deterrents to its commercial use. Health concerns associated with thorium's natural radioactivity have not been a significant factor in switching to alternative nonradioactive materials. US consumption of thorium, all for nonenergy uses, has decreased from 11.4 t (thorium content) to 0.7 t since 1997. The principal applications of thorium today make use of the very high melting point of ThO_2 (3,300°C, the highest of all binary oxides) and of the electron emitting capability of thorium when alloyed with tungsten for use in filaments for high-powered magnetrons for radar.

In the short term, thorium is available for the cost of extraction from rare-earth processing wastes. In the longer term, large resources of thorium are available in known monazite deposits in India, Brazil, China, Malaysia, and Sri Lanka. The world total thorium resources identified and prognosticated amounts to 3.6 million tons Th. Though reported values vary because of the difficulty in measuring such low concentrations, ^{232}Th is present in seawater at only about 0.050 wppb, due primarily to the insoluble nature of its only oxide, ThO_2 . Thus the recovery of thorium from seawater is not a realistic option.

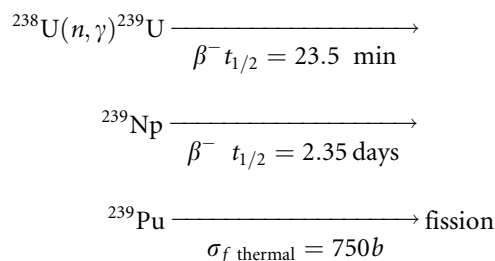
Because ^{232}Th is the only isotope of natural thorium, there are no enrichment plant tails from thorium nuclear fuel. Therefore, the cost of thorium

in a mixed thorium-uranium LWR fuel or in a pure thorium- ^{233}U fuel cycle is relatively small. However, the cost of chemically processing of ThO_2 -based fuel and the separation of ^{233}U is significant.

Energy Content of Uranium and Thorium

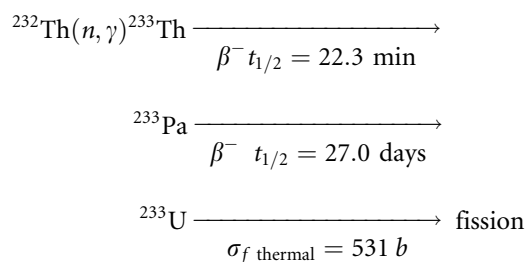
Uranium has 18 known isotopes, none of which are stable and only two of which have half-lives longer than a million years, ^{235}U (704 Ma) and ^{238}U (4.47 billion years). Only ^{235}U , which is about 0.711 wt% (0.720 atom%) of natural uranium, is fissile, i.e., will fission using thermal (i.e., low velocity) neutrons. Uranium-238, which is by far the dominant isotope at 99.2745 wt%, will fission if struck by high-energy neutrons. However, ^{235}U is the only naturally occurring isotope of any element capable of sustaining a neutron chain reaction in a suitably designed reactor.

^{238}U is a fertile isotope and can be transformed into ^{239}Pu through the capture of a neutron and two subsequent beta decays, as shown in the following reaction:



where β^- indicates a beta decay with electron emission and $\sigma_f \text{ thermal}$ is the fission cross section in barns at a neutron energy of 0.025 eV.

Thorium has 25 isotopes, of which only the non-fissile isotope ^{232}Th is long-lived, with a half-life of 14 billion years. However, in the reaction shown below, ^{232}Th can be transmuted into ^{233}U , a fissile isotope:



The fission energy of ^{233}U is 190 MeV and that of ^{239}Pu is 200 MeV. If 1 kg of thorium were bred into ^{233}U , the fission energy available would be $78.9 \times 10^{12} \text{ J}_{\text{thermal}}$ ($78.9 \text{ TJ}_{\text{th}}$). The fission energy in 1 kg of natural uranium, bred to ^{239}Pu , is $80.4 \text{ TJ}_{\text{th}}$. Thus thorium and uranium are quite similar in maximum energy content, but uranium is far more important in the near term because ^{235}U is a naturally occurring fissile isotope.

Global Estimates of Overall Uranium and Thorium Resources

Uranium and thorium have two unique characteristics when compared with other fuels. First, their energy is contained in the nucleus, rather than in the chemical bonds between the atoms, as is the case with fossil fuels. Thus, chemical reactions within the earth, due to pressure, high temperatures, or the presence of oxygen, have no effect on the nuclear energy available from uranium or thorium. In contrast, exposure of fossil fuels to the oxygen in the atmosphere or to volcanic activity releases the energy stored in their chemical bonds. The vast majority of the solar energy originally stored in fossil fuels through photosynthesis of the source biomass has been lost in chemical reactions with the atmosphere, groundwater, and lava.

A second, less obvious, characteristic of uranium and thorium is that they are constantly signaling their presence via the products of radioactive decay. Everyone is familiar with pictures of the prospector with a Geiger counter searching for uranium. Gamma rays and beta particles can be detected with a handheld instrument if the uranium ore is at the surface. However, even as little as a meter of overlying soil will shield the gammas and beta particles from the counter. Therefore, any ore deposits more than a meter below the surface will have to be detected through well logs or core samples or via their gravitational or magnetic signatures, rather than through their radiation.

Astrophysical Origins of Uranium

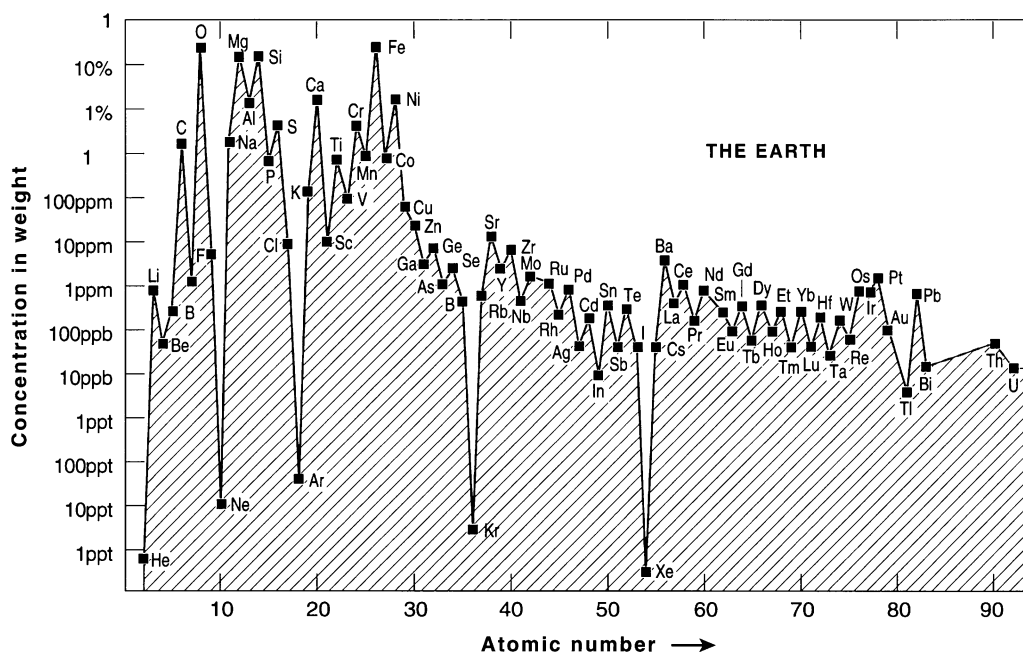
Uranium, thorium, and all other elements heavier than nickel result from the sudden collapse of massive stars as supernovae. The lifetime of stars and the results of these gravity-driven implosions are very dependent on the stars' initial mass. A star having the mass of our sun

lasts for about ten billion years but can only produce elements up to iron. A star having ten solar masses lasts for only 10 Ma until it explodes as a supernova, producing all the elements in the periodic table.

The groundbreaking work by Burbidge, Burbidge, Fowler, and Hoyle [6] led to the realization that all of the elements heavier than nickel are the result of less than a minute of tremendous neutrino and neutron fluxes during the collapse and explosion of a supernova [7]. The nuclide distribution as a function of time in a supernova has been simulated [8] and indicates that isotopes with the maximum number of neutrons ("the neutron drip edge") form during few seconds of intense activity at the center of the imploding supernova. From this nuclear modeling of a supernova explosion, it can be inferred that uranium and thorium are about seven orders of magnitude below silicon in the composition of the supernova debris – the material from which planets are formed. This is in rough agreement with Bulk Silicate Earth model, shown in Fig 1.

Earlier studies are also in agreement. Urey cites estimates by Goldschmidt of the primordial abundance of 41 weight parts per billion (wppb) for uranium and 106 wppb for thorium. Alpher's theoretical curves and Harrison S. Brown's observed astrophysical data show uranium approximately 6.5 orders of magnitude less abundant than silicon, resulting in a primordial abundances of 57 wppb. Deffeyes, accounting for the decay of uranium since the expansion of the primordial neutron gas, estimates global uranium abundance at 10.5 wppb [9].

Recent work on the physics of supernova collapse offers some insight into the expected global inventories of uranium and thorium. These two elements, and all other elements heavier than nickel, are formed in a few seconds of extremely violent conditions during the collapse and explosion of massive stars. During the last few minutes of such a massive star's evolution, hydrogen, helium, and all of the elements lighter than nickel at the center of the star are depleted through fusion reactions. With no more energy available for continued fusion reactions, the center cannot withstand the outer shells of material, and the matter in the center is compressed to a degenerate state in which matter is broken into the constituent particles, primarily neutrons and neutrinos. The torrent of neutrons from the center of the supernova irradiates the infalling



Uranium and Thorium Resources. Figure 1
Composition of the bulk silicate Earth

outer layers of stellar material, producing heavier isotopes at a rate faster than the radioactive decay of those isotopes. The result is the production of isotopes stretching from nickel through uranium and beyond, all saturated with neutrons.

This type of supernova explosion is estimated to occur, somewhere in the universe, at the rate of one per second. Obviously, most such explosions are too distant or masked by dust clouds and are not detected from the earth. Since the beginning of the universe, some interstellar material has gone through multiple cycles of collapse, explosion, dispersal, and accretion into new stars.

The hydrodynamic instabilities of the implosion result in a wide variation in the shapes of the resulting nebulae. Nevertheless, neutron transport and reaction codes have been developed to estimate the distribution of isotopes resulting from a supernova implosion. Wanajo and others [8] have modeled the first few seconds of isotope production and shown that the uranium mass should be about seven orders of magnitude less than that of silicon. Since the chemical and planetary accretion characteristics of silicon, uranium, and thorium are similar, and since the earth is about

10% silicon, one would expect that the overall concentration of uranium in the earth is about 10 wppb. The geoneutrino data from KamLAND and from newer detectors indicate that the global uranium inventory is, in fact, about 10 wppb.

Therefore, based on these astrophysical models, it is fairly clear that the earth taken as uniform body contains about 10 wppb uranium and about 40 wppb thorium. Stated in other terms, the present global inventory is thus 63 Tt (63×10^{12} t) of uranium and approximately 400 Tt of thorium. Although this inventory is a vast amount of both elements, if uranium and thorium had a uniform distribution throughout the earth, as assumed in the cold accretion model, concentrations of uranium and thorium would be far too small to be economically extracted.

Geoneutrino Estimates of Uranium and Thorium

In the last 20 years, however, another decay product of the 4.5 billion year half-life of ^{238}U and the 14.2 billion year half-life of ^{232}Th has been used to estimate the total global inventory of uranium and thorium. These particles, called neutrinos, are extremely difficult

to detect and most neutrinos pass completely through the earth without interacting. Thus neutrino detectors are usually a thousand tons in mass and must be located deep underground to avoid unwanted signals caused by cosmic rays.

Neutrinos occur in three types: electron, muon, and tau. Each of the three types has a corresponding antineutrino. Neutrinos originating within the earth, termed *geoneutrinos*, are actually electron antineutrinos primarily resulting from the decay of ^{40}K , ^{238}U , and ^{232}Th . Geoneutrinos provide a means for estimating the total uranium and thorium content of the earth and also may provide limited information on the location of those resources. These elementary particles have been measured over the past decade by massive detectors in Japan, Canada, and Europe in an effort to differentiate the radiogenic and gravitational components of the total geothermal energy flux through the earth's surface [10]. Neutrino and antineutrino fluxes have also been measured to understand neutrino oscillations, to investigate solar fusion processes, and as a first signal of supernova events. Neutrinos (and antineutrinos) travel close to the speed of light, have a small mass (<2 eV), and lack an electric charge. When an electron antineutrino collides with a proton, the result is a neutron and a positron (i.e., an antielectron). This reaction, known as the neutron inverse β decay, was used in the first detection of the neutrino in the Cowan–Reines experiment of 1956. Following the neutron inverse β decay, the positron reacts with a nearby electron to produce two 511 keV gamma rays. The neutron is absorbed by a hydrogen nucleus, releasing a characteristic 2.2 MeV gamma with a mean delay of ~ 200 μs . Circuitry in the detector registers a neutrino event through the delayed emission of a 2.2 MeV gamma following two 511 keV gammas.

The KamLAND (the Kamioka Large Antineutrino Detector), in central Japan, consists of a 18 m diameter spherical vessel which in turn contains a 13 m diameter nylon balloon. The balloon contains approximately 1,000 t of a liquid scintillator (mineral oil, benzene, and fluorescent compounds). The volume between the balloon and the spherical vessel contains highly purified oil which shields the balloon from external radiation and provides buoyancy to support the liquid scintillator. About 1,900 photomultiplier tubes are mounted on the inner surface of the spherical vessel.

Surrounding the spherical vessel is a water Cherenkov detector which provides additional shielding and acts as a muon veto counter.

The decay chain of ^{238}U into ^{206}Pb results in six antineutrinos, one antineutrino for each beta decay. Similarly, the decay of ^{232}Th in ^{208}Pb results in four antineutrinos [11]. Because the neutron inverse β decay requires an electron antineutrino threshold energy of 1.80 MeV, KamLAND cannot detect ^{40}K antineutrinos, but antineutrinos from both ^{238}U and ^{232}Th are within the range of this instrument.

The overall results of the KamLAND geoneutrino study [12] show that the sum of the global U and Th inventory is approximately 30×10^{16} kg. Since the global Th/U mass ratio is 3.9, the global U inventory is about 6×10^{16} kg or ~ 10 ppb of the mass of the earth. The geoneutrino signal also indicates that the majority of the uranium is in the upper continental crust (UCC) and that relatively little of the inventory is in the oceanic crust, the mantle, or the core. The partitioning of the uranium among the upper, middle, and lower continental crust and the upper mantle occurs via geochemical processes [13].

Thermal models of the earth point to inevitable melting of the earth soon after its accretion due to gravitation energy and due to radioactive decay of uranium, thorium, and potassium. Because of its large ionic size and heating due to radioactive decay, uranium is transferred into low melting temperature fractions and out of the earth's core and mantle into the crust. These geochemical and geophysical models predict that two thirds of initial 63 Tt of uranium present in the earth are now concentrated in the crust, which constitutes only 0.4% of the earth's total mass. The low uranium and high iron concentrations predicted for the earth's mantle and core have been supported by concentrations in iron meteorites and in mantle issuing from oceanic spreading zones (0.1 ppm U), compared with U concentrations in magma and crust in subduction zones, (2 ppm U).

Preliminary results from the newer antineutrino detector Borexino at Gran Sasso in the Apennines [14] generally confirm the KamLAND results but indicate a geoneutrino flux 60% higher. Because of very low radioactive contamination in the materials of construction for Borexino, a signal-to-noise ratio of 50:1 was achieved. This greater sensitivity allowed the Borexino

Uranium and Thorium Resources. Table 1 The main properties of geoneutrinos

Decay	Q (MeV)	$\tau_{1/2}$ (10^9 year)	$E_{\max}$ (MeV)	ϵ_H (W/kg)	$\epsilon_{\bar{\nu}_e}$ ($\text{kg}^{-1} \text{s}^{-1}$)
$^{238}\text{U} \rightarrow ^{206}\text{Pb} + 8\ ^4\text{He} + 6\ e + 6\ \bar{\nu}_e$	51.7	4.47	3.26	0.95×10^{-4}	7.41×10^7
$^{232}\text{Th} \rightarrow ^{208}\text{Pb} + 6\ ^4\text{He} + 4\ e + 4\ \bar{\nu}_e$	42.7	14.0	2.25	0.27×10^{-4}	1.63×10^7
$^{40}\text{K} \rightarrow ^{40}\text{Ca} + e + \bar{\nu}_e$	1.32	1.28	1.31	0.36×10^{-8}	2.69×10^4

$\bar{\nu}_e$ denotes electron antineutrinos

where:

Q is the energy release for the overall decay chain

$\tau_{1/2}$ is the half-life of the parent isotope

$E_{\max}$ is the maximum antineutrino energy in the decay chain

ϵ_H is the heating rate, per kg of the parent isotope

$\epsilon_{\bar{\nu}_e}$ is the electron antineutrino source rate, per kg of the parent isotope

researchers to place an upper bound on the power of any critical fissioning zones in the core at 3 TW, significantly below the indicated global radiogenic heat production of about 18 TW. Collection of geoneutrino data by Borexino is continuing.

Geoneutrino data collected to date indicates that the uranium content of the earth is several orders of magnitude greater than conventional resource estimates. Limited geoneutrino data and an understanding of geochemical processes suggest that most of that uranium content is in the upper continental crust. This data provides some confidence that, with further local exploration or advanced extraction technologies, sufficient uranium could be found for several centuries of expanded nuclear power (Tables 1 and 2).

The overall results of the KamLAND geoneutrino study [10, 12] show that the sum of the U and Th inventory is $3\text{E}17$ kg and since the global Th/U mass ratio is 3.9, the global U inventory is $6\text{E}16$ kg or 10 ppb of the mass of the earth, in agreement with the supernova production ratio with silicon. Further note that the geoneutrino signal indicates that the majority of the uranium is in the upper continental crust and that relatively little of the inventory is in the oceanic crust. The partitioning of the uranium into the UCC via geochemical process is discussed in the next section.

Mechanisms for the Concentration of Uranium

Unlike other energy resources such as coal or petroleum, the resources of uranium are not fundamentally changed by geological processes. Whereas petroleum

Uranium and Thorium Resources. Table 2 U, Th, and K global inventories, radiogenic heating, and neutrino luminosities according to the Bulk Silicate Earth (BSE) model

	m (10^{17} kg)	H_R (10^{12} W)	L_ν (10^{24} s^{-1})
U	0.8	7.6	5.9
Th	3.1	8.5	5.0
^{40}K	0.8	3.3	21.6

might be lost through evaporation or combustion or a natural gas reservoir may vent into the atmosphere, uranium is lost only through radioactive decay or through the relatively rare formation of a natural reactor. Therefore the primordial inventory of uranium, reduced by radioactive decay, remains present somewhere in the earth. The crucial question is “where?”

The natural distribution of elements in the earth's crust is controlled by two major factors. The first is the set of ambient geological fractionating processes that leads to regions of depletion and concentration of the element. The second factor includes the overall geochemical characteristics of the element. Elements that are concentrated by a small number of fractionation processes can be expected to have a multimodal distribution, with a peak in the tonnage versus grade curve for each of the modes of geochemical concentration. For elements having a large number of applicable concentration processes, the peaks overlap and the resulting tonnage versus grade curve takes on

a log-normal characteristic. For example, the element chromium, whose distribution at high concentrations is solely governed by fractional crystallization in mafic magmas (i.e., high in magnesium and iron), one would expect a bimodal distribution of concentrations, with one peak at the average crustal abundance and the high concentration peak at the mafic fractionation concentration. On the other hand, most elements, uranium included, can undergo a wide variety of fractionating processes, and deposits would be expected over a wide range of concentrations. In this latter case, the tonnage versus grade distribution would be expected to be log-normal. Bear in mind that geological conditions change over time and therefore the distribution patterns have varied with time.

In considering uranium in particular, it is important to examine the tectonic and igneous processes that have redistributed the uranium within the crust. In the past four billion years, the most important processes are continental accretion and plate tectonics. In the accretion process, crust formed into masses of continental dimensions. In the second, continuing, process, the continental crust and the oceanic crust have taken on quite different characteristics in terms of uranium concentration.

Igneous Processes

Igneous processes begin with the melting of mantle rocks at depths of 60–200 km, followed by the migration of less dense liquids to the surface. The migration of these less dense minerals to the surface is a predominant process in the formation of the continental crust. The extruded liquid forms crust in two general locations, at mid-oceanic ridges, where the upwelling material forms new oceanic crust and in subduction zones, where the oceanic crust plunges back into the mantle, usually passing under the edge of a continent.

The behavior of uranium in igneous processes is dominated by two characteristics of the element. In the +4 oxidation state, the condition expected in the earth's mantle, the U^{+4} ion has an ionic radius of 97×10^{-12} m (picometers, pm) about the same as Na^{+1} ion (97 pm). Other ions common in the core and mantle are significantly smaller in radius: Fe^{+2} , 74 pm; Ni^{+2} , 69 pm, Mg^{+2} , 66 pm; and Al^{+3} , 51 pm.

Thus, like sodium and the other large ions, uranium ions selectively enter partial melts within the mantle and are transported to the surface.

The second characteristic of uranium is its radioactivity, serving as a source of heat for melting the mantle and core. Like Th^{+4} (ionic radius 102 pm) and K^{+1} (133 pm), these heat-producing elements are readily fractionated out of the mantle and toward the surface. Deffeyes notes that the earth would be a radically different place if the heat-producing elements had small radii, since the geothermal energy source would then be located deep within the core and the convection currents driving plate tectonics would be much stronger [15].

The rocks forming the oceanic crust at mid-oceanic ridges are characterized by a uniform uranium concentration of about 0.1 wppm. Conversely, the crust formed above subduction zones is characterized by uranium concentrations of about 2 wppm. The wide difference in concentration is due to the differences in the source materials and to the different chemistry. The upwelling mantle at the oceanic ridge has a uranium concentration of about 0.005 wppm, while the subduction zones have as their source material oceanic crust and bits of continental crust, with an average uranium concentration of about 0.1 wppm. The continuous upwelling at the oceanic ridges serves as a mechanism for depleting the core and mantle of uranium and incorporating that uranium in the oceanic crust. The relatively low concentration of uranium in the oceanic crust is augmented with uranium from continental runoff, which subsequently precipitates in the ocean basins. At the subduction zones, the oceanic crust is again subjected to partial melting and the uranium is again fractionated in the melt and transported to the surface.

Average Vertical Distribution of Uranium and Thorium

As a result of the various igneous processes, the average concentration of uranium is highest at the surface of the continental crust and decreases approximately exponentially with depth.

The anticipated variation of uranium concentration with depth is given by the equation $U(z) = U(z=0) e^{(-z/h_r)}$, where z is the depth in m, h_r is the depth parameter (discussed below), and $U(z)$ is

Uranium and Thorium Resources. Table 3 Sources of heat in the upper continental crust

Isotope	Thermal output
U-238	0.095 mW/kg
Th-232	0.027 mW/kg
K-40	3.6 nW/kg

the concentration at depth z , in wppm. $U(z = 0)$ is the average continental crustal abundance of uranium at the surface, 2.76 wppm.

This approximation is based on the presence of heat-producing elements, U-238, Th-232, and K-40, in the continental crust, measurements of the thermal conductivity of the crustal materials, and the linear temperature distribution with depth measured at many locations. The heat produced in the crust is divided about evenly between U-238 and Th-232, since the crustal abundance mass ratio between Th and U is 3.9. K-40 is about four orders of magnitudes lower, although potassium has a crustal abundance of 2.1%, since K-40 is only 117 ppm of natural potassium and the thermal energy output of K-40 is about four orders of magnitude below U-238 and Th-232, as shown by Lachenbruch, below [16] [17] (Table 3).

Obviously, this method assumes one-dimensional heat transport and a fairly uniform thermal conductivity, without a significant contribution from flowing fluids. A more recent review by Brady et al. [18] provides more details on the technique.

Several measured values of the depth parameter are listed in Table 4. [19].

If a depth parameter of 8,500 m is assumed, based on the above data, then 11% of the crustal uranium inventory would be expected to be within 1,000 m of the surface and 21% within 2,000 m.

Geochemical Beneficiation Processes

Uranium occurs in ores such as uraninite [UO_2 , pitchblende], carnotite [$\text{K}_2(\text{UO}_2)_2\text{V}_2\text{O}_8 \cdot 3(\text{H}_2\text{O})$], autunite [$\text{Ca}(\text{UO}_2)_2(\text{PO}_4) \cdot 11(\text{H}_2\text{O})$], uranophane [$\text{Ca}(\text{UO}_2)_2(\text{HSiO}_4)_2 \cdot 5\text{H}_2\text{O}$], davidite [$\text{Ce}_{0.75}\text{La}_{0.25}\text{Y}_{0.75}\text{U}_{0.25}\text{Ti}_{15}\text{Fe}_3 + 5\text{O}_3$ and $\text{La}_{0.7}\text{Ce}_{0.2}\text{Ca}_{0.1}\text{Y}_{0.75}\text{U}_{0.25}\text{Ti}_{15}\text{Fe}_3 + 5\text{O}_3$], torbernite [$\text{Cu}(\text{UO}_2)_2(\text{PO}_4)_2 \cdot 12\text{H}_2\text{O}$], and other minerals containing U_3O_8 (actually a stable complex oxide of $\text{U}_2\text{O}_5 \cdot \text{UO}_3$).

Uranium and Thorium Resources. Table 4 Temperature distribution depth parameter

Location	h_r (m)
Sierra Nevada	10,000
Eastern USA	7,500
Norway and Sweden	$7,200 \pm 700$
Eastern Canadian Shield	$7,100 \pm 1,700$
Canadian Appalachians	$10,000 \pm 2,000$
US Appalachians	$8,100 \pm 1,300$

In the other references the depth parameter is denoted as D, rather than h_r .

The governing characteristic in the geochemical transport of uranium is the fact that uranium is highly soluble in oxidizing environments and essentially insoluble in reducing environments. The change in the earth's atmosphere from a reducing to an oxidizing condition about 1.8 billion years ago is thus responsible for a fundamental change in the dominant processes in uranium transport. In the earlier age, igneous processes and fractionation of uranium in partial melts due to its large ionic size were dominant. In the last 1.8 billion years the transport of uranium by means of groundwater oxygenated at the surface has been dominant.

Thus, in the period more than 1.8 billion years ago, uranium was primarily concentrated in placer deposits as a chemically inert and physically dense phase. Because of the low solubility of uranium in reducing environments, rivers, lakes, groundwater, and thus the sea contained very low uranium concentrations. The placer deposits at Elliott Lake, Canada and at Witwatersrand, South Africa are typical of the deposits formed during this period.

With the dominance of photosynthesis in the last 1.8 billion years, the atmospheric and groundwater conditions have been oxidizing and uranium minerals have been highly soluble in the sedimentary weathering cycle. Placer deposits no longer formed and, in fact, began to dissolve. The uranium content of rivers, lakes, and groundwater increased and gradually, the uranium concentration in the oceans also increased. Nevertheless, the uranium concentration remained well below saturation.

In a few isolated locations, however, oxidation of organic-rich beds by groundwater led to locally

reducing conditions. In these locations, the uranium ions or their complexes would reach supersaturation and re-precipitate. An important example of this re-precipitation is in the Mesozoic sandstones of the Colorado Plateau. The uranium ores are found in organic-rich zones where the oxygen in groundwater was removed by carbon-rich debris. Precipitation of uranium has also occurred where restricted circulation in the oceans and organic-rich sediments led to anoxic conditions. Good examples are the black Chatanooga Shale and the phosphorite shale of the Phosphoria Formation [9].

Specific Deposit Types

Before the discovery of the McArthur River deposit, the highest grade uranium ores were obtained from igneous sedimentary deposits such as Great Bear Lake in Canada, Joachimsthal in the Czech Republic, and Katanga in the Congo. The deposits at Oklo in the Gabon Republic in Africa were of high enough concentration and the uranium, 1.7 billion years ago, contained 3% ^{235}U rather than the present 0.711% such that several critical natural reactors occurred in the deposit. The reactor zones released about 15,000 MW-years of fission energy over the course of about 250,000 years. These deposits were formed by the movement of hot water through fractures in blocks of rock heated by their own uranium and thorium content.

Precambrian sandstones overlie much older Precambrian granites and metamorphic rocks. At the interface, there is a discontinuity in the age of the rocks. This type of discontinuity is termed an unconformity. Unconformity deposits, such as those in Saskatchewan and northern Australia occur where uranium from the sandstone, has formed into veins in the open spaces of the interface, and has been heated to temperatures of several hundred degrees Celsius.

Roll-Front Deposits As mentioned earlier, uranium oxide precipitates when the solution enters a reducing environment. The uranium oxide can be redissolved in situ by oxygenated leach solutions. In sandstone deposits, the uranium minerals have been deposited in the interstices between the sand grains. The deposits are often moving very slowly through the sandstone

because of the flow of groundwater, much like the movement of a front through a liquid chromatography column. Oxygenated water from the surface enters the sandstone where reducing agents, such as sulfides or organic matter, are located in the interstitial spaces. The organic carbon in one pore volume of sandstone can remove all the oxygen dissolved in 50,000 pore volumes of oxygenated groundwater.

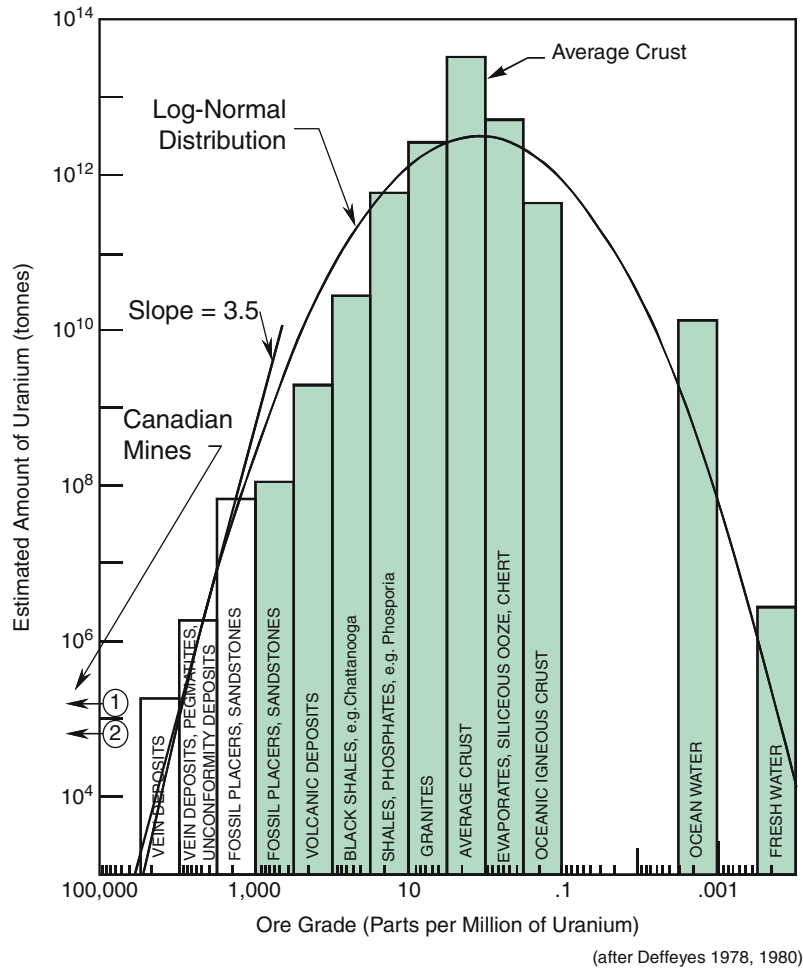
Therefore, the front between the oxygenated groundwater and oxygen-free groundwater moves slowly through the sandstone. Uranium dissolved at the surface and uranium dissolved from the sandstone by the oxygenated groundwater is swept along and precipitated at the front. Upstream of the front the uranium is present in the groundwater as the soluble hexavalent uranyl carbonate complex. As the oxygen is removed from the groundwater at the front, the soluble hexavalent uranium is reduced to the insoluble quadrivalent state.

The quadrivalent uranium precipitates in the form of the mineral uraninite (UO_2). Thus the location of ore bodies is often associated with deposits of carbonaceous materials where the carbon, in much larger quantities compared with the uranium, has removed the oxygen from the groundwater.

Based on the various modes for the formation of uranium ore bodies, reviewing the large body of prior research, Deffeyes and MacGregor estimated the uranium content of the various crustal regimes in a report for the USDOE in 1978 [9].

The distribution of mass versus grade for the various types of uranium deposits is shown in Fig. 2 with additional data [9, 15]. The three gray bars on the left (grade >1,000 ppm U) indicate deposits of the type now mined for uranium alone. The expected log-normal distribution is shown and the mass and grade of two Canadian mines discovered since 1978 are also indicated: (1) the McArthur River deposit, 137,000 t U of proven reserves averaging 18 wt% U and (2) the Cigar Lake deposit, 90,000 t U at an average grade of 17 wt% U. Other known larger, but lower grade deposits, such as Olympic Dam in Australia, have not been shown.

Present mining activities are recovering uranium at market prices of \$150/kg from ores containing 0.027–20% U_3O_8 . Given the log-normal distribution and the known quantities of the various uranium



Uranium and Thorium Resources. Figure 2
Distribution of uranium in the Earth's crust

mineralizations, a tenfold increase in the price of uranium (and thus a tenfold decrease in economically viable ore grade) would result in a 300-fold increase in the amount of uranium available. Equivalently, the World Nuclear Association (formerly the Uranium Institute) estimates that a doubling of uranium prices would result in a tenfold increase in supply [5] (WNA 2010).

The slope of the log-normal curve at presently mined grades is also shown in Fig. 2. This slope, about +3.5, indicates that for a doubling of the cost of mining (i.e., mining ore at half the present concentration), the economically available resources of uranium would increase by more than an order of magnitude ($2^{3.5} \approx 11$). This estimate presumes a continuation of

present mining techniques and does not consider the increased using of in situ leaching and recovery of uranium as a by-product in gold, copper, and phosphate mining.

Unconventional Resources

Existing Enrichment Tails

Another potential source of uranium is the re-enrichment of depleted uranium by using today's over-capacity of enrichment. Because of low price of natural uranium in recent years, many enrichment plants have been operating with tails assays of up to 0.3% ^{235}U . The 1.18 million tons of depleted uranium currently stored at enrichment plants could therefore supplant a few

hundred thousand tons of natural uranium if demand required. The inventory of depleted uranium is expected to increase by about 51,400 t U/year though at least 2010. The enrichment capacity in 2010 was reported to be 57 MWSU/year compared with an annual demand of 49 MWSU/year. The present spare capacity in enrichment plants in the world, around 8 MSWU/year, theoretically represents an equivalent of around 3,000 tons/year of natural uranium if this spare capacity was utilized for enrichment of depleted uranium with an assay of 0.3% and a new tails assay of 0.1%.

The economics of re-enrichment depend on the ^{235}U assay of the depleted uranium and the relationship between the price of uranium and the cost of enrichment services. A tails assay of above 0.3% is preferable if re-enrichment of depleted uranium is to be considered a possibility. Re-enrichment of depleted uranium for the production of low-enriched uranium (LEU) in the Russian Federation has taken place for several years in times of excess enrichment capacity. However, decreasing amounts of excess enrichment capacity makes re-enrichment a marginal source of light water reactor fuel. On the other hand, the eventual use of enrichment tails as breeding blankets for fast reactors, as will be discussed shortly, represents a large long-term source of fuel.

Gold and Phosphate Tailings

In addition to the discovery of new resources through increased exploration, improvements in mining technology are also lowering the cost of previously high-cost deposits. In particular, in situ leaching (ISL) is of growing significance and could be applied to existing gold and phosphates tailings piles. The resource base of 16.2 million tons U does not include uranium in gold and phosphate tailings. The phosphate deposits are estimated at 22 million t U.

Uranium from Seawater

The recovery of uranium from seawater places an upper limit on the cost of uranium. Uranium is dissolved in seawater at 3 mg/m³ (3 wppb) and represents a well-known resource of 4.2 billion tons, 250 times the known land-based resource. The uranium content of the oceans is relatively constant and large-scale extraction can be done without local depletion of the

resource. Since only about 3% of global population lives in landlocked countries, extraction of uranium from seawater is truly the bounding cost for uranium.

New Technologies for Uranium Extraction

The current prices for uranium provide little motivation for the development of new extraction technologies. However, regulations to minimize the impact of mining on the environment and radiation exposure to workers have led to the use of technologies where uranium is extracted in situ or where ores previously mined for another element are processed for uranium extraction. In addition, the technologies described above for the extraction of uranium from seawater could have a major impact in minimizing environmental impact and radiation exposure.

In Situ Leaching

During conventional mining, the rock of the ore body is removed from the ground, transported to a mill, and treated to remove the minerals of economic value. The opening of the mine, the transport of the ore, the milling and the disposal of remaining treated rock can create severe environmental impacts. In situ leaching (ISL), sometimes known as solution mining, involves the use of liquids to dissolve the desired elements from the ore body without removing it from the ground. The liquid is pumped through the ore body and returned to the surface, where the desired elements are removed from the solution by precipitation, by electrochemistry, or other means. The leaching liquid is then returned to the ore body and the process is repeated. ISL eliminates the need to remove large quantities of ore from ground and to transport it to the mill, thus minimizing surface disturbance. ISL also eliminates the need to dispose of the tailings or waste rock. However, for ISL to be effective, the ore body must be permeable to the flow of the leaching liquid. Furthermore, the ISL site must be located so as not to contaminate ground water away from the ore body.

Today, because of its reduced surface impact and lower cost, ISL is used for 85% of US uranium production. Most of the operations in Wyoming, Nebraska, and Texas are less than 10 years old. Worldwide, about 16% of world uranium production uses ISL, including all the production in Uzbekistan and Kazakhstan.

ISL can be used to extract uranium from deposits below the water table in permeable sand or sandstone, provided that the deposit is confined above and below by impermeable strata. Suitable candidates are often roll-front deposits as described earlier. The uranium mineral are usually uranium oxide or uranium silicate coatings on the individual sand grains. The ISL process replicates, in a few months, the conditions that led to the formation of the roll-front deposit in the sandstone initially.

There are two types of ISL, depending on the chemistry of the deposit and groundwater. When the ore body is limestone or gypsum, i.e., containing significant amounts of calcium, then an alkaline leaching agent such as sodium bicarbonate and CO_2 must be used. Otherwise an acid leaching agent, such as weak sulfuric acid plus oxygen at a pH of 2.5–2.8 (about the same as vinegar) is preferred. ISL in Australia is primarily acid, while ISL in the USA is primarily alkaline.

Generally the uranium is extracted by progressively drilling wells into the deposit on a rectangular grid with ~ 30 -m spacing. The leaching fluid is pumped into four wells surrounding a central extraction well, into which a submersible pump has been lowered. The wells are cased to assure that the fluids do not enter strata above the deposit. In the USA the production life of an individual alkaline ISL well is typically 6–10 months. The most successful operations have extracted 80% of the uranium from the ore. Production life is often limited when the sandstone is plugged by mobilized clay and silt. Sometimes the blockages can be dislodged by reversing the flow through the field or by increasing the injection pressure.

The uranium is recovered from the extracted solution in an ion exchange or solvent extraction process. Solvent extraction is preferred if the groundwater is saline, while ion exchanges is most effective if the chloride content is below 3,000 ppm. With alkaline leaching, ion exchange is effective until the total dissolved solids reach 3,000 ppm. The uranium is then stripped from the resin or solution for further processing [20].

Before the process solution is reinjected, it is reoxygenated or recharged with sulfuric acid, for alkaline or acidic processes respectively. About 1% of the process solution is bled off to maintain a pressure gradient toward the wellfield. The pressure gradient

ensures that groundwater from any surrounding aquifer flows into the wellfield and that ISL mining solutions does not enter the aquifer.

Recovery of Uranium from Seawater

The recovery of uranium from seawater is highly speculative and may never prove to be economic. One cubic meter of seawater contains 3 mg of natural uranium, which can deliver 244 MJ_{th} in a breeder or about 2.5 MJ_{th} in a present day LWR. Simple calculations show that the pumping energy needed in an extraction plant could easily consume all the energy available, particularly in the LWR case. Thus seawater extraction conceptual designs relying on ion exchange or adsorption have utilized ocean currents or wave action to move the seawater past the uranium-collecting surfaces.

However, the magnitude of the seawater resource places an upper limit on the cost of uranium for several reasons. First, seawater is available to nearly all countries of the world at virtually the same uranium concentration and without local depletion due to the extraction of uranium. Secondly, because no group of countries can form a cartel over the uranium supply if seawater extraction is practiced, the price of uranium is unlikely to be driven artificially high through market manipulation. Furthermore, the only present limitation on the extraction of uranium from seawater is knowledge of the technology and resins. Thus one would expect that, if conventional sources of uranium become limiting, a healthy competition in research and development would drive down the cost of extraction.

Uranium recovery from seawater has been studied in Japan for a very long term or to face a very strong development of fission energy. In a laboratory scale experiment performed by the Japan Atomic Energy Agency (JAEA) where uranium is trapped by an amidoxime adsorbent which has been prepared on nonwoven polyethylene material with the aid of radiation-induced cograftering. This experiment, 7 km offshore from Sekine-Hama in Aomori Prefecture, Japan, produced more than 1 kg of U on 350 kg of nonwoven fabric during a total submersion time of 240 days [21]. However, at this stage of the study, it is difficult to predict the practical application of uranium recovery from seawater. An economic assessment has

been reported indicating a possible cost for this uranium process in a 1,000-t U/year commercial plant of approximately \$600/kgU [22].

Impact of Uranium Scarcity and Higher Extraction Costs

Table 5 shows the approximate impact of increases in the price of natural uranium on the cost of electricity from a light water reactor. In this set of calculations, the cost of natural uranium is set at \$140/kg U (approx. average price of domestic U to US utilities, 2011), \$500/kg U, representing an optimistic cost of extraction from seawater and \$1,000/kg U, representing a more pessimistic (or perhaps more realistic) cost for extraction from seawater. Two burnups are shown for each uranium price, 45 MW-day (thermal) per kg of initial uranium (MW-day/kg) and 60 MW-day/kg. The specific power of the fuel remains constant at 37.9 kWth/kg of initial uranium, as does the interest charged on the fuel during the fuel cycle. In the cases with higher cost uranium, the tails assay ("Tails U-235 content") to optimize the balance between raw materials and enrichment costs. As shown in the last two lines of the table, an increase in the cost of natural uranium from \$140/kg U to \$1,000/kg U results in an increase in the cost of electricity of about \$0.018/kW-h. For comparison, an increase in the prices of natural gas of \$2.80 per million BTU, as has occurred five times since 2005, also results in an increase in the cost of electricity of \$0.018/kW-h.

Uranium Compared with Future Energy Needs

A simple calculation is needed to place the magnitude of current uranium mining in perspective. If it is assumed that the world population reaches a steady-state level of 10 billion and each of those people consumes energy at the average rate of a US resident in 2011, then the total annual world consumption of energy would be about 3.7×10^{21} J_{th}. While that high rate of consumption would probably not be sustainable for a variety of other reasons, the required natural uranium input to a system of fast reactors to produce 3.7×10^{21} J_{th} would be about 45,000 t U. Average worldwide uranium usage, from both mining and the downblending of HEU, is now about 59,000 t U/year [4].

The Need for Fast Reactors

The early development of fast breeder reactors and of thermal breeder reactors using thorium was driven in part by the apparent global scarcity of fissile isotopes to fuel a rapidly growing set of nuclear reactors. Slow growth in nuclear power and large discoveries of natural uranium in a few regions of the world have reduced the global urgency for breeding more fissile material. In those regions of the world lacking known uranium resources, particularly Europe and Japan, there has been continuing interest in the development of sodium-cooled fast breeder reactors. Such reactors reduce the need for natural uranium by a factor of ~50, compared with light water reactors and thus the cost and availability of natural uranium is a much smaller consideration.

However, beyond the breeding of fissile isotopes, the use of a fast neutron spectrum offers unique capabilities for the consumption of the long-lived actinides, particularly plutonium, neptunium, americium, and curium.

In the long term, the capability of a fast reactor to make use of both ²³⁸U and ²³⁵U will be critical in meeting future energy needs. However, in the next century, fast reactors will be crucial for the management of actinides and the reduction of the long-term radiotoxicity of the nuclear fuel cycle by at least two orders of magnitude.

Comparison of Fossil Fuel and Uranium Reserves

It is interesting to compare the cited uranium known reserves and the inventory of depleted uranium with the estimated reserves of coal, oil, and natural gas. In Table 6, the reserves of fossil fuels and uranium are compared on the basis of known, economically recoverable reserves.

Fossil fuel reserves are those of the World Energy Council, the *Oil & Gas Journal*, and *World Oil* [23]. In the use of Table 6, the caveat cited above applies. Resources of both fossil fuels and of uranium are undoubtedly much larger than the reserves cited. Nevertheless, it is noteworthy that the uranium resources are energetically equivalent to about 20% of natural gas or oil resources, even with the use of LWRs only. The 1.2 million tons of

Uranium and Thorium Resources. Table 5 Impact of uranium costs on the cost of electricity

Impact of uranium costs on the cost of electricity							
	Current uranium prices		Uranium from unconventional sources				Units
			Optimistic: \$500/kg		Pessimistic: \$1,000/kg		
	UO ₂ fuel 45 MWd/kg	UO ₂ fuel 60 MWd/kg	UO ₂ fuel 45 MWd/kg	UO ₂ fuel 60 MWd/kg	UO ₂ fuel 45 MWd/kg	UO ₂ fuel 60 MWd/kg	
Specific power	37.9	37.9	37.9	37.9	37.9	37.9	kWth/kg
Cycle parameters							
Total cycle length	3.6	4.8	3.6	4.8	3.6	4.8	years
Capacity factor	90%	90%	90%	90%	90%	90%	
Effective full power days	1183	1578	1183	1578	1183	1578	efpd
Burnup	45	60	45	60	45	60	MWd/kg ihm
Feed U-235 content	0.711%	0.711%	0.711%	0.711%	0.711%	0.711%	wt%
Product U-235 enrichment	4.00%	5.34%	4.00%	5.34%	4.00%	5.34%	wt%
Tails U-235 content	0.25%	0.25%	0.20%	0.20%	0.10%	0.10%	wt%
Separative work	5.874	8.692	6.546	9.636	8.953	13.020	kg-SWU/ kg fuel
Natural uranium	8.089	10.964	7.439	10.050	6.385	8.569	kg/kg fuel
	0.180	0.183	0.166	0.168	0.142	0.143	kg/MWd
Rates							
Interest rate	5.0%	5.0%	5.0%	5.0%	5.0%	5.0%	/year
Natural uranium	\$140	\$140	\$500	\$500	\$1,000	\$1,000	/kg U
Conversion U3O8 to UF6	\$54	\$54	\$193	\$193	\$385	\$385	/lb U3O8
	\$11	\$11	\$11	\$11	\$11	\$11	/kg natural U
Separative work	\$150	\$150	\$150	\$150	\$150	\$150	/kg-SWU
Costs							
Natural uranium	\$1,132	\$1,535	\$3,719	\$5,025	\$6,385	\$8,569	/kg fuel
Separative work	\$881	\$1,304	\$982	\$1,445	\$1,343	\$1,953	/kg fuel
Conversion	\$89	\$121	\$82	\$111	\$70	\$94	/kg fuel
Fabrication	\$500	\$500	\$500	\$500	\$500	\$500	/kg fuel
Total cost	\$2,602	\$3,459	\$5,283	\$7,081	\$8,298	\$11,116	/kg fuel
Interest during use	\$234	\$415	\$475	\$850	\$747	\$1,334	/kg fuel

Uranium and Thorium Resources. Table 5 (Continued)

Impact of uranium costs on the cost of electricity							
	Current uranium prices		Uranium from unconventional sources				Units
			Optimistic: \$500/kg		Pessimistic: \$1,000/kg		
	UO ₂ fuel 45 MWd/kg	UO ₂ fuel 60 MWd/kg	UO ₂ fuel 45 MWd/kg	UO ₂ fuel 60 MWd/kg	UO ₂ fuel 45 MWd/kg	UO ₂ fuel 60 MWd/kg	
Total fuel cost	\$2,837	\$3,874	\$5,758	\$7,931	\$9,045	\$12,450	/kg fuel
	\$63.19	\$64.73	\$128.27	\$132.50	\$201.48	\$208.00	/MWth-day
	\$0.772	\$0.790	\$1.566	\$1.618	\$2.460	\$2.540	/million BTU
	\$7.81	\$8.00	\$15.86	\$16.38	\$24.91	\$25.72	/MWe-h
COE increase due to high uranium prices			\$8.05	\$8.38	\$16.91	\$17.71	/MWe-h
			0.8	0.84	1.69	1.77	cents/kWe-h

Uranium and Thorium Resources. Table 6 Comparison with fossil reserves

Coal		909	Billion tons
	=	2.01E + 22	Jth
	=	19,089	Quadrillion BTU (quad)
Petroleum		1340	Billion barrels
	=	8.20E + 21	Jth
	=	7,772	Quad
Natural gas		6261	Trillion cubic feet
	=	6.84E + 21	Jth
		6,480	Quad
Total fossil reserves		3.52E + 22	Jth
	=	33,341	Quad
Uranium		5.404	Million t natural U
	+	1.2	Million t depleted U
	=	2.33E + 21	Jth in LWRs
	=	2211	Quad in LWRs
	=	5.36E + 23	Jth in fast reactors
	=	508,000	Quad in fast reactors

depleted uranium currently in storage itself represents an energy source larger than the fossil reserves if used in a fast reactor.

Future Directions

Uranium is ubiquitous in the continental crust and concentrated in economically recoverable deposits by several relatively well-understood processes. Today uranium is being mined from the richest and most convenient of the deposits though little exploration has taken place in the last 20 years. Uranium and thorium are often being extracted as by-products of mining for other elements. It is likely that other similarly rich deposits exist in relatively unexplored regions of Asia and Africa.

Prices in the present uranium market are dominated by large discoveries in the last 20 years and by the conversion of military HEU to civilian purposes. The continued use of nuclear energy and the end of downblending can be expected to raise uranium prices, encourage exploration, and return the uranium to a slightly higher-priced equilibrium.

However, because of the wide range of igneous and geochemical processes that are responsible for the

formation of uranium deposits, it can be expected that uranium will be found in significant quantities with renewed exploration.

Emerging technologies for the extraction of uranium, particular in situ leaching, will make resources in sandstone and shale economically recoverable and minimize the surface disruption due to open pit mining and the occupational radiation exposures of underground mining.

In more distant future, the extraction of uranium from seawater will make this fuel available to virtually every nation. While extraction from seawater is likely to be five to ten times more expensive than uranium is today, the overall increase in the cost of electricity or other energy products would be minimal.

Therefore, the need for fast reactors in the near term, with a global view in mind, is not driven by a scarcity of uranium but rather by a need to effectively manage the long-lived actinides in spent fuel. A fast neutron spectrum is uniquely capable of fissioning the higher actinides and reducing the long-term radiotoxicity and volume of the nuclear fuel cycle. Likewise, although thorium is more abundant than uranium, the primary use of thorium will probably not be for the breeding of ^{233}U , but rather as a host material for the transmutation of the higher actinides in fast neutron spectrum reactors.

There are many challenges in the development of safe, proliferation-resistant, and economical reactors and fuel cycles. Fortunately, the uranium and thorium resources do not appear to be a near-term limitation.

Bibliography

1. Bates RL, Jackson JA, American Geological Institute (eds) (1984) Dictionary of geological terms, 3rd edn. Anchor Books, New York
2. Benedict M, Pigford TH, Levi HW (1981) Nuclear chemical engineering. McGraw-Hill, New York
3. NRC Glossary (2011) Yellowcake. <http://www.nrc.gov/reading-rm/basic-ref/glossary/yellowcake.html>. Updated 27 May 2011
4. OECD Nuclear Energy Agency, International Atomic Energy Agency (2010) Uranium 2009: resources, production and demand: a joint report. OECD, Paris
5. World Nuclear Association (2010) Supply of uranium, Dec 2010. www.world-nuclear.org/info/inf75.html
6. Burbidge EM, Burbidge GR, Fowler WA, Hoyle F (1957) Synthesis of the elements in stars. *Rev Mod Phys* 29(4):547
7. Burrows A (2000) Supernova explosions in the Universe. *Nature* 403:727–733, 17 Feb 2000
8. Wanajo S et al (2003) The r-process in supernova explosions from the collapse of O-Ne-Mg cores. *Astrophys J* 593:968–979, 20 Aug 2003
9. Deffeyes KS, MacGregor ID (1978) Uranium distribution in mined deposits and in the Earth's crust: final report. GJBX-1(79). Dept of Geological and Geophysical Sciences, Princeton University, Princeton. Prepared for the USDOE, Grand Junction Office, Aug 1978
10. Araki T et al (2005) Experimental investigation of geologically produced antineutrinos with KamLAND. *Nature* 436:499–503, 28 Jul 2005
11. Fiorentini G, Lissia M, Mantovani F, Vannucci R (2004) Geo-Neutrinos: a short review. *arXiv:hep-ph/0409152v1*, 14 Sep 2004
12. Enomoto S, Ohtani E, Inoue K, Suzuki A (2005) Neutrino geophysics with KamLAND and future prospects. arxiv.org/ftp/hep-ph/papers/0508/0508049.pdf
13. Herring JS (2004) Uranium and thorium resources. In: Cleveland CJ (ed) *The encyclopedia of energy*. Academic, Amsterdam/Boston
14. Borexino Collaboration (2010) Observation of geo-neutrinos. *Phys Lett B* 687:299–304
15. Deffeyes KS, MacGregor ID (1980) World uranium resources. *Sci Am* 242(1):66–76
16. Lachenbruch AH (1968) Preliminary geothermal model of the Sierra Nevada. *J Geophys Res* 73(22):6977–6989, 15 Nov 1968
17. Lachenbruch AH (1970) Crustal temperature and heat production: implications of the linear heat-flow relation. *J Geophys Res* 75(17):3291–3300, 10 June 1970
18. Brady RJ, Ducea MN, Kidder SB, Saleeby JB (2006) The distribution of radiogenic heat production as a function of depth in the Sierra Nevada Batholith, California. *Lithos* 86:229–244
19. Schubert G, Turcotte DL, Olson P (2001) Mantle convection in the Earth and planets. Cambridge University Press, Cambridge/New York, p 146
20. World Nuclear Association (2010) In situ leach (ISL) mining of uranium. <http://www.world-nuclear.org/info/inf27.html>. Updated Mar 2010
21. Seko N et al (2003) Aquaculture of uranium in seawater by fabric-adsorbent submerged system. *Nucl Technol* 144: 274–278
22. Joint Nuclear Energy Action Plan, Fuel Cycle Technology Working Group (2010) Uranium extraction from seawater, 1st information exchange meeting. Lawrence Berkeley National Laboratory, Berkeley, Meeting record, 24–25 June 2010
23. EIA (2010) Annual energy review 2009, DOE/EIA-0384(2009). Energy Information Administration, US Department of Energy, Washington, DC, Aug 2010

Uranium in the Environment: Behavior and Toxicity

GLEN A. BIRD

EcoMetrix Incorporated, Mississauga, ON, Canada

Article Outline

Glossary

Definition of Subject

Introduction

Nuclear Fuel Cycle

Background U Concentrations

Uranium Speciation

Terrestrial Environment

Aquatic Ecosystems

Uranium Toxicity

Future Directions

Bibliography

Glossary

Acute toxicity test A toxicity test of short duration in relation to the life span of the test organism (e.g., ~ 4 days for fish).

Chronic toxicity test A toxicity test that spans a significant portion of the life span of the test organism (e.g., 10% or more) and examines effects on such parameters as metabolism, growth, reproduction, and survival.

Concentration ratio (CR) The ratio of the steady-state concentration of a substance in an organism to the concentration of the substance in filtered water and/or the ratio of the uptake-rate constant to the depuration-rate constant.

Critical toxicity value (CTV) The quantitative expression of low toxic effect (e.g., EC_{10}) to the measurement endpoint. CTVs are used in risk characterization for the calculation of an Estimated-No-Effect Value (ENEV).

Depleted uranium (DU) The by-product of uranium enrichment and has a ^{235}U content that is less than that of natural uranium (0.71 wt.%) usually 2 to 4 wt.%.

EC_x The concentration of a substance in water, soil, or sediment that is estimated to cause a specified toxic

effect (e.g., immobilization, reduced growth) on $x\%$ (most often 50%) of the test organisms. The duration of the test must be specified. The EC_x describes quantal effects, lethal or sublethal, and is not applicable to quantitative effects (see IC_x).

Estimated-no-effect value (ENEV) The concentration of a contaminant that should not have an effect on a sensitive endpoint of ecological relevance such as survival, growth, or reproduction, and is derived by dividing the critical toxicity value by an appropriate application factor.

IC_x The inhibiting concentration for a substance for a specified $x\%$ impairment in a quantitative biological function such as the attained size over a growth period, reproductive performance, or respiration.

LD_{50} The dose that causes mortality in 50% of the organisms tested.

Lowest-observed-effect level (LOEL) The concentration at which actual ecotoxic effects become apparent, usually the fifth percentile of the screening-level concentration.

No-observed-effect level (NOEL) The highest dose in a toxicity test not causing a statistically significant effect compared with the controls.

Partition coefficient (K_d) A measure of the propensity of a particular radionuclide to associate with solid phases defined as the ratio of the concentration of the radionuclide on the particulate fraction to the concentration in water ($\text{L}\cdot\text{kg}^{-1}$ dry sediment).

Transfer coefficient (TC) A ratio of the concentration ratio (unit less) and the rate of intake of the organism ($\text{kg}\cdot\text{day}^{-1}$) to predict the fraction of the radionuclide in the diet that is accumulated in the body on a daily basis ($\text{day}\cdot\text{kg}^{-1}$).

Definition of Subject

Uranium (U) is a primordial radionuclide that is naturally present in the environment at low concentrations. Natural U is mainly a chemical hazard as opposed to a radiological hazard. The chemical speciation of U influences its bioavailability and toxicity. Under anoxic conditions, U is in the tetravalent state forming insoluble compounds and is immobile. Under oxic conditions, U is in the hexavalent state is mobile, bioavailable, and tends to have higher toxicity. Low absorption in the gut and a tendency to decrease in

concentration up the food chain makes U toxicity of little concern at typical background concentrations, but at elevated concentration U may be toxic.

Considerable work has investigated the toxicity of U to humans using animal models. This work is applicable to wildlife. In comparison to mammals, little effort has focused on the toxicity of U to terrestrial plants and soil invertebrates. For some aquatic species, U toxicity decreases with an increase in hardness and alkalinity. For some species U toxicity is not ameliorated with increasing hardness and alkalinity, and U can be toxic at low concentrations.

This entry provides an overview of the presence of U in the environment, its chemical speciation, behavior in the environment, and toxicity to organisms at low concentrations to assist the practitioner in establishing ecotoxicological benchmarks for protection of the environment.

Introduction

Uranium is a member of the actinide series of elements with an atomic number of 92 and atomic weight of 238.029. Uranium is a heavy, silvery-white, ductile, and slightly paramagnetic metal, which is pyrophoric when finely divided [1]. Uranium is not found in elemental form in nature, but is found in about 155 mineral forms [2]. All 16 isotopes of U are radioactive; ^{238}U (99.27% by weight), ^{235}U (0.72%), and ^{234}U (0.0055%) are the three most common isotopes. Natural materials have ^{238}U : ^{235}U ratios of 137.5 ± 0.5 [3]. Uranium decays primarily by emitting an alpha (α) particle (two neutrons and two protons) from its nucleus [4]. The penetration range of a typical 5 MeV alpha particle is about 4 cm in air and about 50 μm in soft tissue, not enough to penetrate the superficial keratin layer of the human skin [1]. The dominant isotope, ^{238}U , has a physical half-life of 4.5×10^9 years, giving it a very low specific activity ($1.24 \times 10^4 \text{ Bq}\cdot\text{g}^{-1} \text{ U}$). For these reasons, U is primarily a chemical toxicant rather than a radiological toxicant. Uranium-238, ^{235}U ($t_{1/2} = 7.0 \times 10^8$ years) and ^{232}Th ($t_{1/2} = 1 \times 10^{10}$ years) are the long-lived parents of the U, actinium, and thorium decay series.

Major sources of anthropogenic releases of U include U mining/milling and refining processes, combustion of fossil fuels, roasting of sedimentary

rocks for cement production, and the use of phosphate fertilizers. Atmospheric emissions are mainly through the burning of wood, peat, coal, and petroleum, the roasting of rock minerals in the metal extraction, refining, and cement industries, and the incineration of solid wastes.

Uranium is mainly used in the production of electricity by the nuclear industry. Depleted U ($18.97 \text{ g}\cdot\text{cm}^3$) is denser than iron ($7.87 \text{ g}\cdot\text{cm}^3$), has a high melting point ($1,132^\circ\text{C}$), is highly pyrophoric, with a high tensile strength comparable to most steels, is abundant and is relatively cheap. Therefore, depleted U is used for nonnuclear fuel purposes rather than natural U. Depleted U is used for internal guidance devices for missiles, radio emissions shielding, photographic emulsions, as a catalyst in analytical chemistry, photographic toners, pigments and glazes, protective armor on tanks and other vehicles, and for ammunition [1].

Depleted U is a by-product of U enrichment and has a ^{235}U content that is less than that of natural U, for example, about 70% of the ^{235}U is removed [3] so the ^{235}U content is about 0.2–0.4 wt.% ^{235}U isotope. Depleted U is about 42% less radioactive than natural uranium and is indistinguishable from natural uranium in terms of chemotoxicity. A separate discussion of depleted U and its health and environmental effects as a result of military use is beyond the scope of this entry.

Because of the author's greater familiarity with the Canadian situation, examples are largely from Canada. However, the examples should also be relevant to other parts of the world.

Occurrence

Uranium is relatively common in low concentrations in the earth's crust, averaging about $3 \text{ mg}\cdot\text{kg}^{-1}$ [5]. Concentrations are frequently elevated near U mining/milling operations [6, 7]. The relative abundance of U in the earth's crust is comparable to silver, gold, and the light rare earth elements and is more abundant than tin, mercury, and lead. The concentration of U in non-ore grade rocks is as follows: 1–5 $\text{mg}\cdot\text{kg}^{-1}$ igneous rocks; 3–4 $\text{mg}\cdot\text{kg}^{-1}$ shales; 0.5–1.5 $\text{mg}\cdot\text{kg}^{-1}$ sandstones; and 2–5 $\text{mg}\cdot\text{kg}^{-1}$ limestones. Phosphoratic rocks may have 120–300 $\text{mg}\cdot\text{kg}^{-1}$ of U. Shales and other fine-grained

sedimentary rocks have higher concentrations of U than coarser-grained rock, because of the presence of clays and organic material to which U adsorbs.

Uranium is commonly found in association with oxygen in minerals in the tetravalent (4^+) state as uranous hydroxide ($\text{U}(\text{OH})_4$) or in the hexavalent (6^+) state as uranyl ion (UO_2^{2+}). The main uranium minerals are uraninite (pitchblende) and potassium uranyl vanadate (carnotite). U is also found in phosphatic rocks and monzanite sands in commercially extractable concentrations [8].

In air, U easily oxidizes and is mainly found in its oxidized form in nature. Weathering of rocks releases about 27,275–31,325 t of U annually. Natural leaching of U from the rock involves the oxidation and dissolution of the soluble uranyl ion by surface or groundwater. Oxygenation of ores by bacteria such as *Thiobacillus ferrooxidans* in the presence of moisture produces sulfuric acid. Acid conditions enhance the dissolution of U from rocks. Phosphate ion also increases the solubility of uraninite. In rock unaffected by weathering for a million years, ^{238}U , ^{235}U , and ^{230}Th are in secular equilibrium with their decay products (activity ratios are equal to 1). Weathering of the rock causes disequilibrium.

Uranium is mobile in the environment with most U deposits being derived from nearby geological rock formations, through dissolution, mobilization, and recrystallization/redeposition in receptor formations. For example, the U-rich deposits of the Athabasca Basin of northern Saskatchewan originated from dissolution of U from clastic minerals in the Athabasca Group by heated groundwater, followed by precipitation of the U at redox traps situated near peletic zones in the basement rock [9]. Uranium concentrations in the environment (water, sediment, and soil) reflect the concentrations in nearby rock formations and are generally low. The behavior of U in the environment has been extensively studied and the subject of several reviews [5, 10–14].

Nuclear Fuel Cycle

Mining and Milling

In 1990, about 90% of the world's total U production occurred in nine countries: Australia, Canada, Kazakhstan, Namibia, Niger, the Russian Federation,

South Africa, the USA, and Uzbekistan. Underground mining of the ore accounted for about 40% of the world's production in 1996, open-pit mining for 39%, and in situ leaching about 13%. The remaining 8% was recovered as a by-product of other mineral processing [15]. The mining method depends on the location of the ore, local conditions, and the economics. Uranium ores mined commercially may contain as little as 0.04% U or may be highly concentrated, for example, the average U concentration in ore at McArthur River Mine in northern Saskatchewan is about 20%. Recovery of U from the ores in modern milling operations may exceed 97%.

In Canada, the first mining and milling operations for U commenced in 1942 at Port Radium and in the 1950s near Elliot Lake, Ontario, and Beaverlodge Lake, northern Saskatchewan. Depletion of the ore bodies and a decline in the demand for U resulted in these older mines and mills ceasing operation by the mid-1990s or earlier. Today there are three operating mines and three operating mills in the Athabasca basin of northern Saskatchewan. However, with higher prices for U ore, several new mines are scheduled to come on stream. There are also several decommissioned or abandoned mines in the basin.

In other parts of the world, U mining and milling has been going on much longer than in Canada, for example, in Germany U was discovered in 1789 in the Erzgebirge (Ore Mountain) area and abandoned mines and mills are a common legacy of early operations [3].

Uranium mills extract U_3O_8 from crushed, ground ores either by an acid or by an alkaline leaching process. After leaching, the resultant solution containing U goes through a solvent extraction process in which the U is purified and concentrated. Yellowcake, the end product of the milling process, is shipped to a U refinery. The residual material (tailings), which is crushed ore minus most of the U, is pumped as slurry to a tailings management facility. Uranium tailings contain up to 85% of the radioactivity initially present in the U ore, since most of the U decay products remain in the tailings [16]. Releases from early mines and mills were not treated. Effluent treatment in Canada only started in the 1970s with the introduction of federal regulations. For this reason, the more significant environmental effects of U mining and milling are mainly associated with older operations. At older operations, untreated

tailings were often discharged to a nearby lake which served as a tailings pond. Subsequently tailings were flushed downstream with discharge events contaminating downstream water bodies. An example is the deposition of tailings in Mudford Lake resulted in tailings moving downstream to Langley Bay, Lake Athabasca, covering the bay with a 2–4 m thick blanket of tailings [17].

Uranium mill tailings have been dewatered and deposited directly on land. These are relatively dry and prone to wind erosion. Dust particles contaminated with U and decay products can be dispersed over much larger areas. Dust particles from tailings and from mill vents constitute a potential exposure pathway for terrestrial biota through inhalation. Wet and dry deposition may increase U concentrations in surrounding soils and vegetation [18] and hence U enters the terrestrial food chain albeit in a localized area. For example, the impact due to aerial transport of aerosols and radon generated from mining in Brazil was restricted to an area less than 15 km² [15].

Tailings may also be a source of U released to groundwater [6]. Uranium enters surface waters as a result of releases of treated mill effluent and from mine dewatering activities. Generally, the effluent treatment plants also treat contaminated water collected from tailings management facilities.

Uranium operations not only release U to the environment but they release other contaminants including members of the U decay chain such as ²²⁶Ra, ²²²Rn, ²³²Th, ²¹⁰Pb, and ²¹⁰Po; metals such as cadmium, copper, bismuth, arsenic, and molybdenum; organic contaminants such as kerosene, amine, and isodecanol which may be released when the mill's solvent extraction process is upset [19]; and salinity stressors [6].

Refining, Conversion, Enrichment, and Fabrication Facilities

The major processing steps involved in U refining and conversion are the purification of yellowcake from U mills to UO₃ in the UO₃ circuit, and conversion of UO₃ to either uranium hexafluoride (UF₆) or uranium dioxide (UO₂). Yellowcake is chemically complex, but contains 26–28% ammonium and magnesium

diuranate, which are biologically available. Yellowcake is refined by nitric acid digestion to an intermediate product, uranyl nitrate, which is thermally decomposed into uranium trioxide (UO₃). This in turn is either converted to uranium dioxide (UO₂) as an end product, or reacted with hydrofluoric acid to form uranium tetrafluoride (UF₄), which is calcined and further fluorinated to produce uranium hexafluoride (UF₆). The UO₂ is pressed and sintered into high-density fuel pellets, and inserted into zircalloy tubes assembled in bundles to serve as fuel in heavy-water reactors that use natural uranium.

UF₆ is one of the few non-uranyl compounds of hexavalent uranium. UF₆ is volatile at temperatures above 60°C, is highly corrosive, and is stored as uranium metal. Enrichment plants take UF₆ from a conversion facility and enrich it usually to 3–5 wt.% ²³⁵U depending on the specific reactor requirements. An end product of enrichment is tailings of depleted uranium (DU) ranging from 0.25 to 0.35 wt.% ²³⁵U [20]. Re-enrichment is the further depletion or stripping of depleted uranium tailings to produce U with a ²³⁵U content of natural U (0.71%). Thus the tailing is upgraded to natural ²³⁵U grade, which may then be further enriched to reactor grade [21]. Enriched uranium hexafluoride (3–5% ²³⁵U) is used as reactor fuel in power reactors or as weapon grade material (>90% ²³⁵U). The enriched UF₆ is converted into uranium dioxide powder that is then processed into pellet form. The pellets are then fired in a high-temperature, sintering furnace to create hard, ceramic pellets of enriched uranium. Releases that occur at different stages of the U refining and conversion process are vented to the atmosphere or treated and released to surface waters.

Nuclear Power Plants

Nuclear power plants generally do not release significant quantities of U and, therefore, there is no U impact on the environment from these sources. Impacts on the environment from nuclear power plants are mainly confined to impingement, entrainment, and thermal effects on aquatic ecosystems. Releases of radionuclides to the environment are generally low and, although of great concern to the public, usually have no effect on the environment or to the public.

Waste Management (Nuclear Power Plants)

Waste management facilities for radioactive wastes were developed to handle a variety of wastes ranging from contaminated soils (historical contamination), liquid and solid wastes, contaminated filters and resins from pollution control systems of nuclear power plants and reactor spent fuel. Modern facilities are well-engineered structures with well-controlled releases, whereas older facilities (e.g., pre-1970s) are generally not as well designed and are more likely to have releases that have the potential to impact on the environment.

Background U Concentrations

Atmospheric U concentrations measured over the North Atlantic are about $4 \times 10^{-6} \mu\text{g}\cdot\text{m}^{-3}$ and are considered to represent background concentrations [8]. Typical atmospheric U concentrations range from 0.025 to $0.1 \times 10^{-3} \mu\text{g}\cdot\text{m}^{-3}$ [10]. Higher values are found in developed regions, for example, values from 0.10×10^{-3} to $1.47 \times 10^{-3} \mu\text{g}\cdot\text{m}^{-3}$ are reported over upper New York reflecting industrial emissions [8]. Background concentrations of U in southern Ontario are about $0.1 \times 10^{-3} \mu\text{g}\cdot\text{m}^{-3}$ [22].

In mineralized soil, U concentrations are generally about $1\text{--}4 \text{ mg}\cdot\text{kg}^{-1}$. Fine-textured soils develop primarily from the weathering of sedimentary rocks and have U concentrations similar to the rock ($\sim 4 \text{ mg}\cdot\text{kg}^{-1}$). Uranium concentrations up to $250 \text{ mg}\cdot\text{kg}^{-1}$ may be present in phosphate rock, which points to the importance of phosphorus fertilizers as a source of radiation in the environment (see Waller this section). Uranium concentrations in soil generally increase with increasing clay and organic content, because of the high affinity of U for clay and organic material. Coarser sandy soil has low U concentrations.

Background concentrations of U in surface water including seawater are generally quite low with values usually a few parts per billion or less. The total U concentration in the Atlantic and Pacific oceans is about $3.1 \mu\text{g}\cdot\text{L}^{-1}$ [23, 24]. This natural background concentration in seawater is 9–11 times greater than those for freshwater. Average U concentrations in rivers are $0.2\text{--}0.6 \mu\text{g}\cdot\text{L}^{-1}$ [3]. Concentrations in drinking water in the USA average about $3 \mu\text{g}\cdot\text{L}^{-1}$ with a range of $0.1\text{--}973 \mu\text{g}\cdot\text{L}^{-1}$ [8]. Higher values are probably associated with the use of groundwater as drinking water.

Background concentrations of U in lake water across Canada are generally low with a median concentration of $< 0.05 \mu\text{g}\cdot\text{L}^{-1}$ and a maximum concentration of $1,350 \mu\text{g}\cdot\text{L}^{-1}$ [25]. The maximum value is associated with the U deposits of northern Saskatchewan. The median U concentration in streams is higher at $0.06 \mu\text{g}\cdot\text{L}^{-1}$, with a maximum of $255 \mu\text{g}\cdot\text{L}^{-1}$ in a stream in the Yukon Territory. In northern Saskatchewan, an area known for its U mines, the median aqueous background U concentration is below the detection limit of $0.05 \mu\text{g}\cdot\text{L}^{-1}$ with an upper 95th percentile of $0.35 \mu\text{g}\cdot\text{L}^{-1}$.

Uranium concentrations in sediments in lakes generally range from 0.5 to $5 \mu\text{g}\cdot\text{g}^{-1}$ (dry weight sediment unless specified otherwise), with an average of $3 \mu\text{g}\cdot\text{g}^{-1}$ [5]. Freshwater sediments from Canadian Rocky Mountain streams have U concentrations ranging from 0.1 to $430.0 \mu\text{g}\cdot\text{g}^{-1}$ with a median value of $3.4 \mu\text{g}\cdot\text{g}^{-1}$, whereas U sediment concentrations in Canadian Shield lakes range from 0.1 to $733 \mu\text{g}\cdot\text{g}^{-1}$ with a median value of $4.6 \mu\text{g}\cdot\text{g}^{-1}$. Sediment concentrations in northern Saskatchewan have a geometric mean of $3.7 \text{ mg}\cdot\text{kg}^{-1}$ sediment, with a 95% confidence level (CL) of $29.5 \mu\text{g}\cdot\text{g}^{-1}$ dry weight (dw) sediment. A maximum sediment U concentration of $648 \mu\text{g}\cdot\text{g}^{-1}$ dw and several other high values in northern Saskatchewan were associated with the ore deposits [6]. The few lakes with high U sediment concentrations have been drained and mined.

U concentrations are generally low but under natural conditions can be extremely high in certain areas. Groundwater U concentrations are generally in the range of $0.1\text{--}50 \mu\text{g}\cdot\text{L}^{-1}$ and occasionally as high as $2,000 \mu\text{g}\cdot\text{L}^{-1}$ [11]. Uranium concentrations in groundwater in the USA, Europe, and Australia range from 0.1 to $1,870 \mu\text{g}\cdot\text{L}^{-1}$ with most values less than $2.2 \mu\text{g}\cdot\text{L}^{-1}$ [8]. In the region of Helsinki, Finland, U concentrations in deep groundwater are up to $14,870 \mu\text{g}\cdot\text{L}^{-1}$ [26].

Uranium Speciation

Freshwater

An understanding of U speciation is essential to understanding the bioavailability and toxicity of U. Uranium exists in four valence states: U^{3+} , U^{4+} , U^{5+} , and U^{6+} , but only tetra- and hexavalent states are commonly

found in the environment. Tetravalent U occurs in reducing environments such as anoxic sediments and water, and in mineral formations. The tetravalent form has a high affinity for organic materials and very low solubility ($0.06 \mu\text{g}\cdot\text{L}^{-1}$) [11]. Hexavalent U exists in oxidizing environments such as surface waters and weathered rock [14]. Most mobilization of U occurs during weathering processes in which uranous (U^{4+}) compounds, generally uraninite, are released from rock and oxidized to the uranyl ion by surface waters. The uranyl ion is the principal form of uranium found in aquatic environments. Uranyl ion is a very large divalent positively charged ionic species similar in size to potassium, calcium, and cesium [14]. In natural waters, U usually forms complexes, typically with chloride, fluoride, phosphate, sulfate and carbonate, hydroxide, and perhaps silicate ions and organic materials, with fluoride complexes predominating below pH 5 and phosphate complexes predominant between pH 4 and 7.5 [27]. The formation of U complexes generally greatly increases the solubility of U in oxidizing environments.

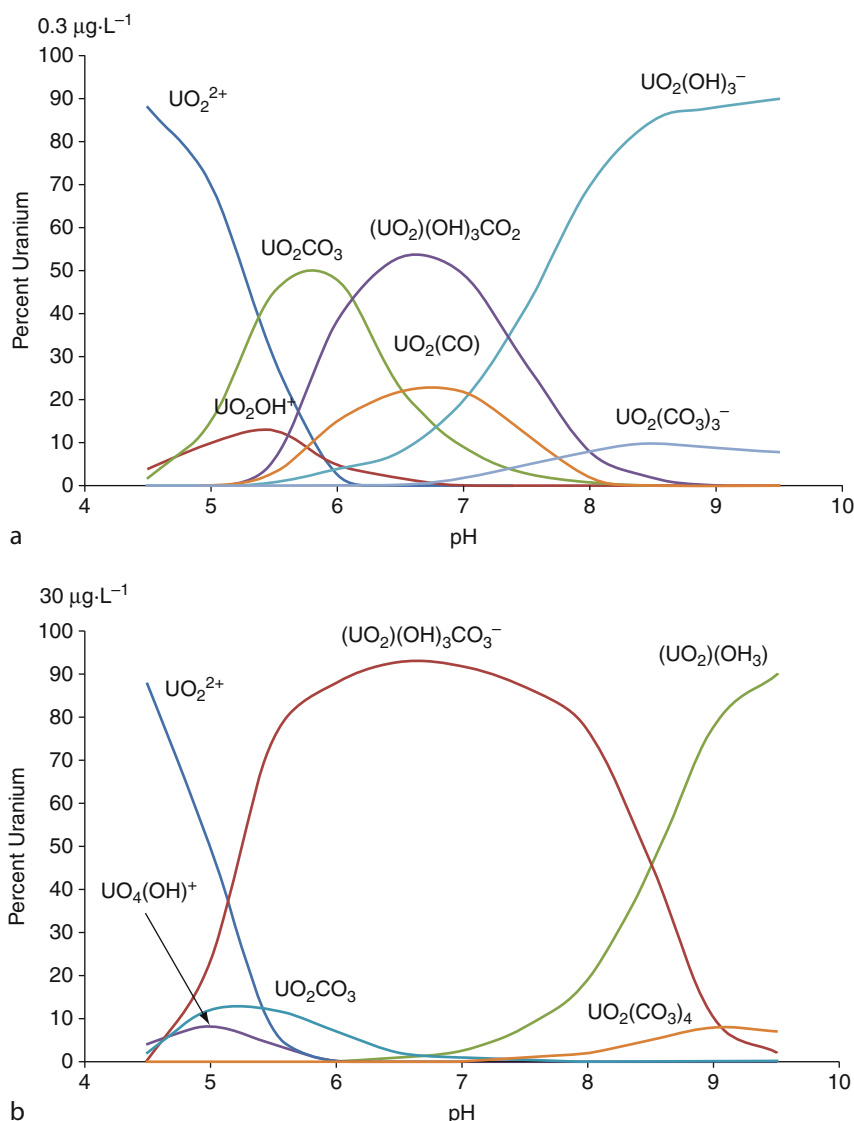
The chemical speciation of U, like that of other metals, is influenced by water quality variables such as alkalinity, hardness, pH, and natural organic matter (e.g., [28, 29]). The pH of a solution has a strong influence on the chemistry of U. U oxides and their hydrolysis products have the highest solubility at pH 4–6. Figure 1 shows the distribution of U species with pH based on thermodynamic simulations. Note that there is large uncertainty in the predictive power of the simulations [7, 30] and that predictions can only be validated for waters with low total dissolved solid concentrations [5]. Furthermore, the output of different geochemical speciation models can differ substantially [31].

Speciation modeling provides evidence that UO_2^{2+} is the most toxic form of U, although some studies suggest exceptions at different pH values [29, 32, 33]. Markich et al. [29] also attribute U toxicity to UO_2OH^+ . UO_2SO_4 and UO_2CO_3 do not contribute to toxicity [34]. The free uranyl ion (UO_2^{2+}) is calculated to be the predominant species at $\text{pH} \leq 5$, but is insignificant at $\text{pH} > 6$ (Fig. 1). UO_2OH^+ is also important at $\text{pH} \leq 5$. At high pH, the uranyl ion forms soluble complexes with carbonate ions. Uranium carbonate complexes such as uranyl carbonate (UO_2CO_3) are very soluble especially at pH 6 and

higher, and have an overall negative charge. Since CO_2 and carbonates are abundant and readily available in water, uranyl carbonates commonly contribute significantly to U concentrations in water. From pH 5.5 to 6, UO_2CO_3 is the predominant uranyl species. From pH 6 to 7.5, the mixed uranyl-hydroxide-carbonate species, $(\text{UO}_2)_2(\text{OH})_3\text{CO}_3^-$, is the dominant uranyl species, whereas $\text{UO}_2(\text{OH})_3^-$ dominates at $\text{pH} > 7.5$. At high U concentrations ($30 \mu\text{g}\cdot\text{L}^{-1}$), $(\text{UO}_2)_2(\text{OH})_3\text{CO}_3^-$ dominates from pH 5 to 8.5 (Fig. 1b). Polymeric uranyl-hydroxide species are only important in natural waters when U concentrations are greater than $200 \mu\text{g}\cdot\text{L}^{-1}$ [5]. Conditions that favor the formation of free ion UO_2^{2+} include low pH, low natural organic matter concentrations, and low alkalinity [29, 35].

The distribution of U species varies with the total U concentration as well as the concentration of ligands [29, 36]. When the concentration of ligands is much greater than that of the metal concentration, the relative distribution of the metal speciation tends to remain stable. However, when the metal concentration is higher than the ligand concentration, there is a significant shift in the metal distribution with U tending to form hydroxyl and carbonate complexes [32].

The organic carbon of surface water is usually in the range of $1\text{--}10 \text{ mg}\cdot\text{L}^{-1}$ [37]. In brown-water (organic-rich) lakes and streams of low hardness and alkalinity, U complexes predominantly with humic substances. Complexation of U increases up to about pH 6, because of increasing ionization of the carboxylate (COOH) functional groups of humic substances. Uranium is strongly absorbed by the carboxyl group, but is not reduced. The adsorption equilibrium of humic acids is described by the Langmuir iso-pH curves. Above pH 6, U complexation with humic substances decreases, because of the higher binding affinity with carbonate and hydroxide ($\text{UO}_2(\text{OH})_3^-$, $\text{UO}_2(\text{CO}_3)_3^{4+}$, $\text{UO}_2(\text{CO}_3)_2^{2+}$, and/or $(\text{UO}_2)_2(\text{OH})_3\text{CO}_3^-$) [5]. The binding of U to humics is often underestimated in studies, because humic substances are not accounted for in many speciation models, the chemistry of humic substances is poorly defined, their stability constants are poorly known, and most speciation modeling is done with simple, chemically defined experimental water rather than natural water [5]. In most instances, ligands decrease the bioavailability of U [32].



Uranium in the Environment: Behavior and Toxicity. Figure 1

Percent distribution of uranium speciation with pH (4.5–9.5) in freshwater at uranium concentrations of (a) 0.3 µg·L⁻¹ and (b) 30 µg·L⁻¹ (Modified from Markich [5])

Uranium uptake generally correlates with free uranyl ion concentration as predicted by the free ion activity model and biotic ligand model, although hydrophilic metal ligand complexes may increase the bioavailability of U above the predicted free ion concentration [5, 32]. Increased toxicity in the presence of ligands involves organic ligands that can be assimilated by the organism, for example, low molecular weight metabolites such as citrate and thiosulfate.

The ligand is assimilated along with the metal. This is seen in the plant–soil system, where uptake of U by roots can be enhanced by the presence of organic chelators that increase the solubility of U, that is, citric acid addition decreases pH which increases U desorption from the soil and the binding of uranyl by the citrate anion further increases the solubility and mobility of U, increasing U availability for plant uptake [14].

Seawater

In oxic seawater, pH 8, U behaves conservatively and is predominately (80–90%) present as a stable, soluble uranyl tricarbonate complex $\text{UO}_2(\text{CO}_3)_3^{4-}$. Minor portions of U are bound to peroxide ligand, $\text{UO}_2(\text{CO}_3)_2(\text{HO}_2)^{3-}$ [5].

Effect of Mine and Mill Effluents on Speciation

Bernhard et al. [38] assessed uranium speciation in three different uranium mining areas in Germany by measuring speciation using time-resolved laser-induced fluorescence spectroscopy and laser-induced photoacoustic spectroscopy and by speciation modeling. The water samples were filtered through a 1-nm cellulose filter before analysis and did not appear to be treated to remove contaminants or neutralize pH. In mine water from Schlemma, with neutral pH (pH 7.13, total organic carbon of $62 \text{ mg}\cdot\text{L}^{-1}$ and $1.9 \text{ g}\cdot\text{L}^{-1} \text{ SO}_4^{2-}$), the calcium uranyl carbonate complex dominated. In the Königstein mine, U was leached underground with dilute sulfuric acid. This resulted in a very acidic mine water with pH 2.6 and sulfate as the major anion ($2.2 \text{ g}\cdot\text{L}^{-1}$). In the Königstein mine water, uranyl sulfate complexes, mainly $\text{UO}_2\text{SO}_4/\text{UO}_2(\text{SO}_4)_2^{2-}$ and UO_2^{2+} , were the major species. In contrast, the Helmsdorf mill tailing water had high concentrations of sulfate, hydrogen carbonate/carbonate and chloride ions, pH 9.76, and elevated concentrations of phosphate, arsenate, and total organic carbon. In the Helmsdorf tailing water, uranyl carbonate complexes, mainly $\text{UO}_2(\text{CO}_3)_3^{4-}$ and $\text{UO}_2(\text{CO}_3)_2^{2-}$ were dominant. Therefore, water chemistry has a marked effect on U speciation.

At U mines/mills, the release of treated effluent complicates the interpretation of U toxicity to aquatic organisms, since treated effluent not only increases U and sulfate concentrations, but also increases calcium and magnesium concentrations in the receiving surface water. For example, as a result of mill effluent releases in Island Lake from the Cluff Lake Operations (now decommissioned) in northern Saskatchewan, mean baseline hardness increased from $36 \text{ mg}\cdot\text{L}^{-1} \text{ CaCO}_3$ to $1,113 \text{ mg}\cdot\text{L}^{-1} \text{ CaCO}_3$ between 1978 and 1988; calcium increased from 7.18 to $292 \text{ mg}\cdot\text{L}^{-1}$; and magnesium increased from 4.4 to $96.5 \text{ mg}\cdot\text{L}^{-1}$. In comparison, HCO_3^- only increased from 36.3 to

$55.0 \text{ mg}\cdot\text{L}^{-1}$, and CO_3 from about 0.0 to $6.7 \text{ mg}\cdot\text{L}^{-1}$, whereas SO_4 increased from 1.28 to $1,140 \text{ mg}\cdot\text{L}^{-1}$ [39].

To assess the potential compounding effects of effluent treatment on increased U concentrations, U chemical speciation modeling using the PHREEQC code and WATEQ4F database, was performed by Environment Canada and Health Canada (EC and HC) [6] for Beaverlodge Lake, a decommissioned site which used an alkaline leaching process, and for Key Lake Operations. For Beaverlodge Lake carbonate species of U were predicted to represent 98% or more of the total U, with less than 0.5% of U present in the form of the toxic species UO_2^{2+} and UO_2OH^+ . For the Key Lake Operation the presence of elevated sulfate concentrations in the treated effluent resulted in a significant increase in the proportion of UO_2SO_4 (50–60%) and in the toxic U species UO_2^{2+} (17–29%) and UO_2OH^+ (4–17%) and in a decrease in the proportion of carbonate species of U to $\leq 15\%$ as well as the absolute amount. Based on these studies, treated mill effluent from mining and milling operations may increase the concentration of U in the receiving environment including the concentration of toxic U species (UO_2^{2+}). However, it also increases the concentration of calcium and magnesium which may have ameliorating effects on U toxicity. It is obvious from these studies that site-specific studies are required to quantify impacts.

Terrestrial Environment

Atmospheric emissions may contaminate the surrounding terrestrial environment through atmospheric deposition of U. This may be in the form of both dry deposition, where particles and gases impinge directly onto surfaces such as foliage and soils, and wet deposition, or washout, where the particles and gases are extracted from the atmosphere by rain or snow and are deposited with the precipitation.

Dry deposition tends to dominate closer to the source because particles gradually settle out downwind of the plume. At greater distances, the particles remaining in the plume are smaller and tend to stay aloft until precipitation occurs. The extent of dry deposition depends more on features such as topography and surface roughness than does wet deposition. Elevated (hilltop) areas may receive relatively more dry deposition because the landscape effectively intersects

the airborne plume. Particles in the plume become impinged on the elevated surfaces. Surface roughness influences the extent of dry deposition, for example, trees intersect and entrain more airborne particles than low-lying vegetation such as grass.

Wet deposition is less dependent on the topography and surface roughness than is dry deposition. However, the underlying surfaces affect the fate of the contaminants washed out of the plume. Some contaminants may dissolve in the water droplets, whereas others will remain as particles. Dissolution will also happen as the precipitation interacts with contaminants previously deposited on surfaces by dry deposition processes. Dissolved contaminants may be absorbed directly into the foliage or will be washed off the foliage surface. Contaminants can be washed off the surface at the drip line of the vegetation (through fall) or along stems and trunks (stem flow). Through fall and stem flow can result in an uneven distribution of contaminants in the underlying soil with elevated soil concentrations at the drip line and immediately adjacent to the trunk. The wash off is a mass action process that generally follows first-order kinetics with a half-time in the order of 2 weeks regardless of particle size [40].

In the soil, dissolved contaminants will interact with the soil solids. Contaminants accumulate in soil litter, but ultimately the litter is decomposed and incorporated into the underlying soil. Contaminants that sorb to the litter reach the underlying soil in a relatively short time (months to years). Contaminated particles may be incorporated into the soil or be subjected to dissolution and leached from the soil.

Uranium moves in solution in soil both vertically and horizontally. Downward movement is in response to gravitational potential, whereas upward movement is by capillary rise or forces exerted by plant roots extracting water. Horizontal movement or through flow occurs in a sloping landscape coupled with gravitational potential. Redistribution of fine soil particles over time, also contributes to the movement of U in soil. The "A" horizon of the soil profile usually has the highest U concentration, whereas the "C" horizon has the next highest and the lowest is in the "B" horizon. This is because of the leaching and gravity settling of soluble and ultrafine particulate materials from the middle zone to the lower horizon, and because of retention of organically bound U in the surface humus [14].

Mobility of U in soil is a function of soil type and conditions. In the absence of carbonates and under oxidizing conditions, U in soil water is present as an uranyl ion or an uranyl hydrolysis product. Uranyl ion is common at pH 5.5 and below, and has a positive charge. Uranyl hydrolysis products exist above pH 5.5, with either positive or neutral charge. In the presence of carbonates at neutral pH, or above, negatively charged uranyl carbonates are common.

Soil particles and associated organic matter have negatively charged absorption sites to which positively charged metals and complexes can bind. Therefore, positively charged uranous ions, uranyl ion, and uranyl hydrolysis products are relatively immobile in soils, and negatively charged uranyl carbonates are highly mobile [8]. Alluvial soils, which tend to have high clay and organic matter content, bind U tightly, whereas in sandy soil, U is more mobile and are at lower concentrations. Anaerobic soils, such as wet marshy soils or sediments, have reducing conditions, which results in the adsorption of positively charged uranous ions. Under oxidizing conditions U is readily lost from the water to the peat by cation exchange, identical to the exchange of Ca^{2+} and K^{+} and other cations in soil. Therefore, adsorption of U is dependent on cation-exchange capacity, but does not always correlate with the cation-exchange capacity. In a soil with a high organic matter and clay content, mean extractable U was about 10% with a range of 4–24%. In comparison, mean extractable U in a sandy, acidic soil was 40% with a range of 7–65% [41]. Therefore, a relatively large portion of the U associated with soil may be bioavailable (extractable).

Storm events selectively erode finer clay and silt particles that have high contaminant concentrations such as U [42] transporting U downslope from contaminated floodplains to adjacent streams [43] or to topographically low areas such as wetlands and surface water bodies. These areas become secondary locations for exposure.

Direct exposure of terrestrial biota to atmospheric contaminants is from either inhalation or foliar uptake. However, more significant impacts usually occur as a result of accumulation of the contaminants in environmental media. Contaminants that are not very mobile in soil tend to accumulate where they are deposited, which exposes resident biota to elevated concentrations.

Terrestrial plants are exposed to contaminated soil by way of their roots. In the case of trees, their fine roots occupy the area just below the litter layer and tend to be more prevalent at the edge of the tree drip line in order to optimize extraction of nutrients and water. Tree roots will also penetrate unsaturated soil to intersect subsoil moisture, and this may predispose them to absorb contaminants in groundwater. Plant roots have the ability to change the chemical and physical makeup of the soil, to optimize growing conditions. The rhizosphere, which is the area occupied by the roots of plants, has an altered pH, an active microbial community, and may be enriched in organic ligands, some originating in the root, which solubilize elements that would otherwise be insoluble. These special features of the rhizosphere optimize the absorption of relatively insoluble nutrients such as phosphorus, iron, manganese, and zinc and can also influence the bio-availability of contaminants such as U.

Uranium is absorbed with water by the root with U penetrating the epidermis of the root. However, U is precipitated as a yellowish deposit in the meristem region with little U entering the root sap system. The storage of U as a yellow deposit in the cell nuclei of the meristem results in the destruction of chromatin and cessation of the cell nucleus activity, preventing U translocation [14]. However, others have reported that the action of U in plant tissues appears to be mainly at the cell surface. In the shrub, *Corprosma australis*, at least 50% of the U was bound to cell wall proteins in the form of U-protein complexes [8]. The binding of positively charged U complexes to cell walls and membranes and to carboxyl groups of proteins may restrict transport mechanisms responsible for movement and deposition of sugars and other metabolic products in plant cells.

Uranium concentrations are highest in the root, and generally follow the order: roots > fruits (seeds) > branches > bark > leaves (needles) > wood and tend to be higher in older branches [18].

Decomposers such as insects and earthworms recycle nutrients (U) in dead plants and animals. In contaminated areas, plant litter often accumulates on the soil surface because the decomposer community has been disrupted [44, 45]. This not only illustrates the importance of decomposers, but suggests that decomposers may be more sensitive to contaminants than primary producers.

Decomposers such as insects and earthworms are important in the comminution of the litter into smaller pieces, which accelerates decomposition by microbes. Microbes colonizing litter take up contaminants from the surrounding pore water, and transfer the contaminants to higher trophic levels through the detrital food web.

Soil invertebrates are also exposed to contaminants through contact with and ingestion of soil and soil pore water. Soil nematodes live in the pore water, so direct absorption of contaminants in pore water is a primary exposure pathway for nematodes. Larger invertebrates consume soil along with the litter, so their exposure may be directly with soil solids. Feces egested by invertebrates may also be re-ingested by other invertebrates to extract additional nutrients and associated contaminants.

Concentration Ratios, K_d , and Transfer Factors

The migration of contaminants in the environment is complex. However, simplified models have been widely adopted to predict contaminant transfer and the resultant concentration in media. Concentration ratios (CRs), solid/liquid partition coefficients (K_d), and transfer coefficients (TCs) assume that the transfer of contaminants is dependent on concentration. Their values are empirically derived so they can be robust and self-validating.

The CR may also include the assumption that contaminant uptake is for a specified soil depth, usually the depth of cultivation and that uptake is from the root, that is, absorption by foliage is lumped with root uptake. It is also assumed that CR data can be transferred from one setting to another. Databases are available to describe the distribution of CR, K_d , and TC [46–52]. However, it is best to undertake studies to establish site-specific parameter values when possible as generic values may not necessarily describe local conditions [53].

The U content of plants is derived mainly from soil water. Plant to soil concentration ratios generally range from 0.0001 to 0.1 which results in low U concentrations. However, some plant species may have a U concentration as high as 0.5% dw. The mosses, *Sorbus aucuparia*, and *Astragalus* sp., are indicator plants for U and have been used in prospecting for

U ore. *Sphagnum* spp. and *Ledum groenlandicum* reflect soil U concentrations in lowland areas, *Umbilicaria* spp. and *Cladonia* spp in rock outcrops, and *Picea mariana* and *Betula papyrifera* in upland locations [54].

Agricultural plants typically have higher U concentrations. Root crops and leafy crops contain more U than other crops. Starchy roots contain approximately 1 $\mu\text{g}\cdot\text{kg}^{-1}$, cereals 0.5 $\mu\text{g}\cdot\text{kg}^{-1}$, and fruit and vegetables 0.8 $\mu\text{g}\cdot\text{kg}^{-1}$ U [8]. Plants high in potassium and calcium have lower U concentration. Agricultural plants accumulate more U from irrigation waters than from the soil. Concentration ratios between agricultural plants and soils are generally 10^{-5} to 10^{-3} and between plants and irrigation water are 1–100 [14].

As the soil U concentration increases, the U concentration in plants generally increases. An exception is marshy soil where U concentrations may be high, but associated plant U concentrations are low because U is tightly bound under anaerobic conditions. Hence, it is both the amount of U in the soil and soil conditions that determine the plant U concentration. Uranium concentrations in plants in arid areas may exceed that in soil. At neutral to alkaline pH in aerobic soils in arid areas, soluble and mobile uranyl carbonates are available for uptake. There are also plant species-specific differences in uptake. Among different plant species growing in soil with a given U content, U concentration may vary by a factor of 10 [14].

Terrestrial Environmental Pathways/Food Chain

Uranium does not biomagnify and is transferred inefficiently through the food web. U has a high partition coefficient (K_d) in most soils and a relatively low mean plant/soil CR of 0.0045. Uranium concentrations tend to decrease with successive trophic level (Table 1). Therefore, the risk of U toxicity in the terrestrial environment is low.

Uranium is taken up through ingestion of drinking water and contaminated food, inhalation and absorption through the skin. The respiratory route is not usually important except for small mammals living in heavily contaminated areas [56].

Soil Ingestion

Soil ingestion may represent a major source of U for wildlife [57]. Soil ingestion is usually considered an important pathway in environmental risk assessments, but is generally not addressed in feeding studies. Linsalata et al. [58] estimated that over 90% of the daily U intake in cattle from natural sources is from soil. Animals may consume soil incidentally, while ingesting roots and vegetation or while grooming, or may intentionally ingest soil at salt licks that provide nutrients during the spring [59]. Sheppard [40] suggested a mean ingestion rate of 50 $\text{mg soil}\cdot\text{kg}^{-1}$ food in the diet of animals for modeling purposes. Wildlife preying on soil-dwelling invertebrates may ingest a higher proportion. High rates of soil ingestion up to 87% of the diet have been reported for some wild

Uranium in the Environment: Behavior and Toxicity. Table 1 Concentrations of uranium in a plant–rodent food chain and through a plant–deer food chain based on Mahon [55]

Plant–rodent food chain		Plant–deer food chain	
Taxa	ng U·g ^{−1} dw	Taxa	ng U·g ^{−1} dw
Deer mice (<i>Peromyscus maniculatus</i>)	<5	Mule deer (<i>Odocoileus hemionus</i>) flesh	24
Chipmunk (<i>Eutamias amoenus</i>)	10	Mule deer (<i>Odocoileus hemionus</i>) bone	9
Fireweed seed heads (<i>Epilobium angustifolia</i>)	22.1 ± 8.1	Mule deer (<i>Odocoileus hemionus</i>) liver	5
Grouseberry (<i>Vaccinium scoparium</i>)	16.2 ± 10.4	Grouseberry (<i>Vaccinium scoparium</i>)	16.2 ± 10.4
		Lichens (<i>Bryoria freemontia</i> and <i>Alectoria sarmentosa</i>)	17.5 ± 10.4
		Grass (<i>Calamagrostis rubescens</i>)	69.3 ± 71.9

species [57, 59]. Feces of whitetail deer average 29% inorganic matter. Soil ingestion is highly variable and may change seasonally [57].

The amount of U retained from ingested soil is a function of the total amount of U ingested and its bioavailability [40], which varies with soil type. At present, there is no reliable method to predict the bioavailability of U through this exposure route. A gut absorption rate of 5% is often used for adult mammals, with a higher value of 10% for juvenile and immature animals. A value of 30% may be expected for neonates, based on the work of Sullivan [60] for plutonium uptake by neonate swine. EC and HC [6] assumed soil to be 2–5% of the diet depending on the wildlife species, and made the assumption that soil U was assimilated at the same rate as U in food.

Aquatic Ecosystems

Since U in industrial effluent releases is mainly to surface waters, it is the aquatic system that is most affected by effluent releases. In the case of U mining/milling, releases of treated effluents are frequently to small headwater lakes or streams.

The fate of U and other radionuclides will depend on their chemical characteristics and the nature of the water body. The physical state of the contaminant, particulate or dissolved, the pH and chemical composition of the water, particularly dissolved organic matter (DOM) or ligand content will have a strong influence on the fate of the contaminant. In carbonate-rich water, U exists as the uranyl-hydroxycarbonate and is poorly sorbed [37]. In soft-water lakes carbonate concentrations are low, and sorption of the uranyl ion occurs to particulates. Upon entering a water body, contaminants partition among the various compartments. A portion becomes associated with dissolved inorganic and organic ligands in solution. The remainder becomes associated with particulate matter as a result of adsorption, precipitation, coprecipitation, and uptake by plankton and eventually settles out of the water column.

Effluents from mines/mills usually enter the epilimnion of the receiving water body as opposed to deeper waters. Mixing in the epilimnion is rapid. Hesslein [61] found that complete mixing takes place within 24 h in a thermally stratified 16 ha lake. Particle-reactive

radionuclides, such as ^{60}Co , ^{65}Zn , and ^{210}Pb (and U) are rapidly adsorbed to DOM, suspended solids, and epilimnetic bottom sediments or are taken up by algae. Dissolved radionuclides also diffuse into interstitial sediment pore waters.

Continuous input of U mine effluents to the surface (epilimnion) waters of a thermally stratified lake in summer will result in the rapid loss of U to littoral sediments. Sediments in deeper waters will gradually become contaminated through the process of sediment focusing. Despite losses from littoral sediments to deep waters, the U concentration of littoral sediment will not decrease because of the continuous input of U and other contaminants to the lake from the effluent. Therefore, sediment U concentrations in shallow water sediments can be high especially if the sediment is high in organic content.

In the water column, dissolved radionuclides are rapidly taken up by algae. The contaminated algae are fed upon by zooplankton, which in turn is a source of food for higher trophic levels. Particle-reactive radionuclides taken up by algae may be routed to the hypolimnion of the lake via the rapid sedimentation of zooplankton fecal pellets. This appears to be the case for U as algae, sediment-trap material, and deepwater organic particles were found to have similar U (and Ra) concentrations [62]. Uranium associates with organic particles (plants, algae, and microbes) in the water column, which eventually sink to the bottom to produce organic sediments. Sedimentation of algae and aggregates formed by the coagulation of colloidal particles will also result in the transfer of radionuclides to bottom sediments, but at a slower rate than that by fecal pellets. Uranium in sediment may be almost completely associated with algae, debris, and humic acids [63].

The transfer of contaminants such as U from water to sediment is an important process for their removal from the water column. The behavior of contaminants in lake water can be described as being nonreactive (conservative) or reactive (non-conservative), depending upon their sorption to suspended particles and retention within the lake [64, 65]. Nonreactive radionuclides tend to remain in the water column and are lost from the lake via the outflow. Thus, there is little retention of these radionuclides within the lake. In comparison, reactive radionuclides are rapidly transferred to sediment and are retained in the lake.

Bird et al. [65] defined probability distribution functions for the transfer rate of U and other elements from the water column to the sediments based on published data and mass balance equations. A high transfer rate indicates that the element has a high affinity for sorption to particles and is lost from the water column [66, 67]. In lakes, nonreactive radionuclides tend to have transfer rates <0.1 , whereas more reactive radionuclides usually have transfer rates >1.0 . In terms of residence time in a non-flushing system, a radionuclide with a transfer rate of 0.1 would have a half-time in the water column of 6.93 years, whereas a radionuclide with a value of 1.0 would have a half-time of only 0.69 years. The transfer rate of U from water to sediment was found to have a lognormal distribution with a geometric mean of 0.5 and a geometric standard deviation of 3.0. Therefore, the overall behavior of U in the range of conditions expected in lakes essentially falls midway between a particle-reactive element and a nonreactive element and can be considered a moderately reactive radionuclide. U tends to partition to sediments as evidenced by a relatively high K_d of $0.36\text{--}3.2 \times 10^3 \text{ L}\cdot\text{kg}^{-1}$ [68]. K_d values vary among sites and are lower at contaminated sites than uncontaminated sites, that is, as contaminant concentrations increase, K_d values and CRs decrease. This is probably because at high levels of contamination, the number of available sorption sites in the sediments becomes limiting.

In a study of the effect of sediment type, temperature, and natural colloids (humic and fulvic acids) on the transfer of radionuclides from water to natural sediments, Bird and Evenden [69] found that the loss of U from water was generally in the order: organic sediment $>$ clay $>$ sand sediment and was slower for U than cobalt, zinc, and cesium. The transfer rate for U was $0.001\text{--}0.084\text{-day}^{-1}$ depending on the experimental conditions. The transfer rate for U to sediment increased significantly for each 10°C increase in temperature between 5°C and 25°C in the presence of organic sediment, presumably due to enhanced microbial activity. There was no difference with temperature in the presence of sand sediment, whereas a greater loss rate was observed at 5°C than 15°C for the clay sediment. An increase in temperature is reported to reduce the sorption rate of uranyl ions to clays [70].

Loss of radionuclides from water was slower in the presence of natural colloids [69]. The chemical behavior of metals and radionuclides becomes dominated by the physical behavior of the colloids following surface complexation [71]. Colloids are usually negatively charged and are known to increase the mobility of radionuclides [37, 72] and to alter their bioavailability and toxicity [73].

In lakes, size-selective sorting and ultimate transport of smaller particles from littoral areas to the deeper profundal waters, a process termed sediment focusing [74] will, over time, result in greater accumulation of sediment and sediment-associated contaminants in the deeper waters. In general, there is a linear relationship between sediment accumulation and lake depth, with flocculent sediments tending to accumulate at depths greater than 10 m in Canadian Shield lakes [75]. Thus, there will be a gradual transfer of the particle-bound radionuclides from littoral regions to the deeper profundal region via sediment focusing, the result of water mixing events over time. In shallow lakes, sediment focusing is not an important process as sediments are continuously redistributed throughout the basin. In these shallow, well-mixed lakes, little difference in sediment contaminant concentration is expected with water depth, but concentrations will vary with sediment type.

Considerable decomposition of organic particles may occur en route to the bottom sediment, particularly in the metalimnion, resulting in a substantial release of the radionuclides, including U into the deeper waters. In Chesapeake Bay, U is transferred from the water column to the sediments with the detritus that forms during periods of high productivity. Degradation of detritus results in the release of U to bottom waters and porewater and there is redox-dependent cycling of U between sediment and bottom water. During summer, anoxic conditions in the hypolimnion inhibit the exchange of U with overlying water due to the stability of reduced U(IV). However, with reoxygenation of the water with fall turnover, U is released back into the water column [76]. In a meromictic lake which has anoxic bottom waters, U in sediment was mainly associated with colloids (80%) with about 40% associated with fulvic and hydrophilic acids [77].

Once U is transferred to the sediment, decomposition may release U and other contaminants to porewater. Decomposition of algae and other detritus in sediment results in about 20% of the U being released to the porewater. Epiphytic algae or periphyton growing at the sediment–water interface are very efficient at taking up radionuclides released from the sediments and from the overlying water. Bioaccumulation of the radionuclides by periphyton may result in the transfer to higher trophic levels via the food chain [73].

As redox potential (E_h) decreases (usually with depth in the sediment), U(VI) may be reduced to U(IV). The reduction of U(VI) to U(IV) and the subsequent chemical reactions and transformation of U sulfide, carbonate, or other biogenic ore requires strongly reducing conditions, that is, U(VI) reduces to U(IV) at low E_H (<200 mV). In the sediment profile, redox conditions change with oxygen being reduced first, followed by denitrification, then by UO_2CO_3 reduction to uraninite and sulfate reduction to sulfide. Uranium reduction and removal from pore water is related to bacterially mediated sulfide reduction. When the sediment undergoes bioturbation or when sulfate reduction is halted, U is released back into the pore water. The release of U is associated with the reduction of U hydroxides of Fe and possibly Mn [7].

The uranyl ion adsorbs by base exchange onto clay and other crystal layer-lattice minerals in the sediment. Sorption to clay minerals occurs at or below pH 5 and sorption to Fe and Al (oxy)hydroxides, silica, and biotic surfaces occurs at higher pH. Both processes reduce the mobility of U in freshwater. The configuration of uranyl ion (UO_2^{2+}) with the oxygen atoms located collinearly allows for a strong structure when absorbed to minerals. This provides considerable resistance to mobility particularly when adsorbed between the layers of carbonaceous material. The enrichment or depletion of radionuclides, including U is influenced by oxides of iron and manganese. These oxides may reduce U from U^{6+} to U^{4+} . The partitioning of U in sediment can be very stable, for example, Edgington et al. [78] reported that U in sediments of Lake Baikal has not been significantly redistributed over at least the past 250,000 years.

In sediments, trace metals and radionuclides can bind to sediments through absorption at particle surfaces, incorporation into the lattices of primary or secondary minerals, or occlusion in amorphous

material [70, 79, 80]. The binding of metals and radionuclides to sediment reduces their mobility and bioavailability. Jenne [81] proposed that hydrous oxides of manganese and iron are primarily responsible for the fixation of metals to sediment. The binding of metals to the oxides is affected by E_h , pH, concentration of the metal and competing metals, concentrations of other ions capable of forming inorganic complexes and organic chelates [81]. Sediment characteristics, such as particle size, organic content, cation-exchange capacity, and microbial dynamics also influence retention in sediment [70, 79]. Because of internal lake mixing, sediments in profundal areas tend to be finer and higher in organic content than shallow littoral sediments and in many respects similar to the upper most A_o horizon of soils [82]. U concentrations in sediment are dependent on sediment organic matter content [83]. Consequently, U in floodplain soils is largely associated with organic matter as a result of sediment deposition [84] and is associated with the acid-soluble fraction in sequential extractions [85].

Changes in environmental conditions may facilitate the release of radionuclides from sediment to the overlying water. For instance, changing environmental conditions such as E_h , pH, temperature, salinity, biological activity, and ligand types and concentrations may affect the speciation of radionuclides, their complexation, and consequently their mobility, bioavailability and toxicity. If the bottom waters of the hypolimnion are anoxic, U strongly sorbs to particles and the sediment, whereas other redox-sensitive radionuclides may be released back into the water where they persist, mainly in the colloidal form, until reaeration of the water column with lake turnover. Lake mixing will result in the redistribution of the radionuclides throughout the water column and the subsequent loss of particle-reactive radionuclides to bottom sediments [73, 86]. Resuspension of contaminated sediment particles may also result in release of radionuclides into the water column with the degradation of the suspended particles. For example, Simpson et al. [87] found that resuspension of sediments for several hours reduced acid volatile sulfide concentrations, which resulted in release of a significant fraction of the trace radionuclide sulfide phases in the marine environment.

Aquatic organisms may take up radionuclides directly from the water column via respiration, by

adsorption of radionuclides to their outer body surface and from ingestion of food. The latter includes ingestion of suspended particles by filter feeders such as zooplankton and clams, ingestion of bottom sediment by benthic organisms, and the indirect ingestion of sediment in the gut of prey organisms. Benthic organisms are also exposed to radionuclides in sediment pore waters. The major route of exposure may vary with the life stage of the organism and the radionuclide's concentration and bioavailability in a particular medium. For example, when aqueous concentrations are high, the major route for uptake may be from the water, whereas when aqueous concentrations are low, uptake may be mainly from sediments by way of the food chain [73].

Aquatic Food Chain

Uranium does not biomagnify through the food web and tends to decrease with successive trophic levels (Table 2). Therefore, U generally poses minimal risk

to aquatic and semiaquatic wildlife. The exception is waterbodies where U concentrations are greatly elevated.

In contaminated Beaverlodge Lake, the food chain pathway was sediment → insects → forage fish → lake whitefish (*Coregonus clupeaformis*) and white sucker (*Catostomus commersoni*). The bottom-feeding fish, the whitefish, and sucker, had much higher U concentrations than pelagic fishes, for example, lake trout (*Salvelinus namaycush*) [68].

Since U mine/mill effluent releases are mainly to surface waters, it is the aquatic system that is most affected by effluent releases. In freshwater systems, the CR between aquatic plants and water may be high [47], which can lead to secondary toxicity via the accumulation of U in organisms feeding on aquatic plants. Bird and Schwartz [47] reported geometric mean U concentration ratios for shield lakes of 1,780 L·kg⁻¹ (GSD = 1.6) for algae, 1,750 L·kg⁻¹ (GSD of 2.2) for macrophytes, 950 L·kg⁻¹ (GSD = 16) for freshwater invertebrates, and 6.2 L·kg⁻¹ (GSD = 6.2) for freshwater

Uranium in the Environment: Behavior and Toxicity. Table 2 Uranium concentrations in the aquatic food chain

Lake Michigan [88]		Lake in central interior of British Columbia [55]		Beaverlodge Lake Area, northern Saskatchewan [68, 89]	
Taxa	ng U·g ⁻¹	Taxa/media	μg U·g ⁻¹	Taxa/media	μg U·g ⁻¹ ww
Pisivorous fish	0.5–0.7	Water	0.34 ± 0.17 μg·L ⁻¹	White sucker (<i>Catostomus commersoni</i>)	0.06 ± 0.01
Planktivorous fish	0.3–0.6	Rainbow trout (<i>Oncorhynchus mykiss</i>)	<0.005	White fish (<i>Coregonus clupeaformis</i>)	0.08 ± 0.02
Benthic invertebrates	2–8	Finescale sucker (<i>Catostomus Catastomus</i>)	0.005	Lake trout (<i>Salvelinus namaycush</i>)	0.005
Zooplankton	16–44	Snail (<i>Pisidium</i> sp.)	0.4	Spottail shiner (<i>Notropis hudsonius</i>)	1.03
Benthic filamentous algae	51	Plankton	0.6 ± 0.2	Lake chub (<i>Couesius plumbeus</i>)	0.62 ± 0.11
Phytoplankton	88–156	Algae	13.4	Juvenile white sucker (<i>Catostomus commersoni</i>)	2.67
Macrophytes	62–615	Sediment	10.3 ± 6.2	Blackflies (<i>Simulium</i> spp.)	2.6 ± 0.5
				<i>Chironomus</i> spp.	15
				Caddisflies (<i>Nemotaulius</i> sp.)	25.6 ± 2.6
				Sediment	109 ± 22

fish. Concentration ratios tend to be higher at low contaminant concentrations, and lower at higher contaminant concentrations. For example, in lakes near Canadian U mines where U concentrations were elevated, EC and HC [6] calculated site-specific geometric mean concentration ratio of 89 (maximum of 158) was calculated for algae, 1.5 (maximum 38) for macrophytes, and 1.24 (maximum of 38) for fish. Wildlife that feed in and along the contaminated waterways have the potential to become contaminated through ingestion of high levels of U in plants, water, and sediment. EC and HC [6] concluded that at some mine sites consumption of macrophytes, fish, water, and sediment were important pathways for U uptake and risk quotients (RQs) indicated potential toxicity to terrestrial mammals. For example, EC and HC [6] reported a RQ of 0.17 for muskrat at background concentrations of U ($0.35 \mu\text{g}\cdot\text{L}^{-1}$) and RQs up to 21 for muskrat at a site where the U concentration was $1,061 \mu\text{g}\cdot\text{L}^{-1}$. In the marine environment, water exposure was the main route of uptake of U by crab (*Pachygrapsus laevimanus*) and zebra winkle (*Austrocochlea constricta*) [90].

Sediment Ingestion

Sediment ingestion may represent a major source of U for wildlife [57] and is usually considered an important pathway in environmental risk assessments. The amount of U retained from dietary sediment is a function of the total amount of U ingested and its bioavailability (see soil ingestion). Sediment ingestion is highly variable and may change seasonally [57]. Wildlife preying on invertebrates in sediments may ingest much sediment. For example, the diet of mallard ducks on average contained 3.3% sediment but 10% of the 88 specimens examined consumed 26% sediment in their diet at certain times of the year [57]. Sandpipers consumed 7–30% sediment in their diet and two wood ducks contained >70% sediment in their gut. Information is not available on the portion of sediment in the diet of other animals such as the muskrat that often dig into the sediment to acquire their food and feed primarily on roots during winter [91], or the moose which has been observed to feed on the rhizomes of aquatic plants shortly after ice-out in spring, before new plant growth has developed (personal observation). It is

assumed that these animals periodically ingest a high proportion of sediment in their diet. Higher accumulation of U occurs in the roots of cattails (*Typha* sp.) than other parts of the plants [92]. Muskrats and similar animals may also consume sediment indirectly while grooming themselves. Sequential extraction studies to determine the bioavailability of sediment-bound U at Beaverlodge Lake area showed that between 69% and 90% of the U in contaminated sediment may be bioavailable [93]. EC and HC [6] assumed sediment to be 2–5% of the diet depending on the wildlife species and made the assumption that sediment U was assimilated at the same rate as U ingested from food.

Uranium Toxicity

The remainder of this entry focuses on the toxicity of U, with emphasis on more sensitive organisms to assist the practitioner in establishing toxicity benchmarks.

The most commonly used compounds to assess the toxicity of U are uranyl nitrate ($\text{UO}_2(\text{NO}_3)_2\cdot 6\text{H}_2\text{O}$) and uranyl sulfate ($\text{UO}_2\text{SO}_4\cdot 3\text{H}_2\text{O}$). Uranyl acetate ($\text{UO}_2(\text{C}_2\text{H}_3\text{O}_2)_2\cdot 2\text{H}_2\text{O}$) is occasionally used. These ionize in water to form hexavalent uranyl ions (UO_2^{2+}).

Toxicity Benchmark Values

Benchmark values are the measurement endpoint of the environmental assessment. The values are used to protect the species or communities of concern based on toxicological data. Benchmarks are the toxicological concentrations of a contaminant where low effects or no effects are expected to the species, population, or community based on ecologically relevant endpoints such as survival, growth, and reproduction. Application factors are often applied to the lowest observed effect concentration (LOEC) or critical toxicity value (CTV) to ensure protection of more sensitive species than the test species.

Toxicity to Terrestrial Plants

There have been surprisingly few studies on the toxicity of U to plants (Table 3), particularly in comparison to the work on U toxicity to mammals (below). In a literature review, Sheppard [97] reported U toxicity to plants at soil concentrations ranging from 0.5 to $2,600 \text{ mg}\cdot\text{kg}^{-1}$ dw. Work by Aery and Jain [94] showed that the

Uranium in the Environment: Behavior and Toxicity. Table 3 Uranium toxicity to terrestrial plants and soil organisms based on soil concentrations (mg U·kg⁻¹ dry soil)

Taxa	Endpoint	Effect concentration	Reference
Plants			
Wheat (<i>Triticum aestivum</i>) L. var. WH-147	EC ₄₄ # seeds/spike	1	Aery and Jain [94]
	EC ₅₈ # seeds/spike	5	
Cotton and sugar beet	EC ₂₅ for yield	5	Shevchenko et al. (1987 cited in [94])
Wheat (<i>Triticum aestivum</i>), and <i>Lycopersicon esculentum</i>	EC ₂₅ for biomass yield	6	Gulati et al. [95]
Native range grasses purple threeawn (<i>Aristida purpurea</i>), buffalograss (<i>Buchloe dactyloides</i>) and little bluestem (<i>Schizachyrium scoparium</i>)	LOEC for biomass increase	50	Meyer et al. [96]
Crop plant	EC ₁₁ for yield	50	Zhukov and Zudilkin (1971 cited in [97])
Scots pine (<i>Pinus sylvestris</i>)	NOEC for growth	>100	Sheppard et al. [98]
Rice	EC ₂₅ for survival at harvest	250	Bufatin et al. (1985 cited in [99, 100])
Crop plants and pine (<i>Phaseolus vulgaris</i> , <i>Zea mays</i> , <i>Lactuca sativa</i> , <i>Lycopersicon esculentum</i> , <i>Brassica rapa</i> , and <i>Pinus strobes</i>)	EC ₂₅ for pod yield in beans	300	Sheppard et al. [99]
Grass (<i>Elymus lanceolatus</i>)	NOEC for emergence, length, and biomass	1,000	Sheppard et al. [100]
Soil fauna			
Collembola (<i>Onychiurus folsomi</i>)	EC ₂₀ for survival and reproduction in fine sand	92–190	Sheppard et al. [100]
Earthworm (<i>Lumbricus</i> sp.)	LOEC for survival	300	Sheppard and Evenden [100]
Collembola (<i>Folsomia candida</i>)	EC ₂₀ for survival in fine sand	350	Sheppard et al. [100]
Collembola (<i>Onychiurus folsomi</i>)	EC ₂₀ for survival and reproduction in loam soil	>390	Sheppard et al. [100]
Collembola (<i>Folsomia candida</i>)	EC ₂₀ for survival and reproduction in loam soil	>710	Sheppard et al. [100]
Earthworm (<i>Lumbricus</i> sp.)	NOEC for survival	1,000	Sheppard et al. [100]

addition of only 1 mg U·kg⁻¹ dry soil affected the growth of wheat (*Triticum aestivum*). At 1 mg U·kg⁻¹, there was an 11% decrease in the number of spikes per plant, 44% decrease in the number of seeds per spike, and 8% decrease in seed weight. This is a toxic effect to a plant species at a U concentration representative of background conditions. A dose-dependent response

was seen in each of these endpoints as soil U concentration increased. At 125 mg U·kg⁻¹ soil, number of spikes and seeds per spike each declined by more than 70%. Thus, U may inhibit the growth of certain plants at natural concentrations. In contrast, Sheppard et al. [99] found that seed germination of corn, lettuce, tomato, rapeseed (*Brassica rapa*), and

pine was not affected by U at concentrations as high as 1,000 mg U·kg⁻¹ dw in garden soil (high organic carbon content), presumably because of the lower bioavailability of U. In limed brunisol soil, seed germination was significantly reduced at 1,000 mg U·kg⁻¹ dw. Pine seeds exposed to 1,000 mg U·kg⁻¹ dw soil did not germinate, whereas seed germination was not affected by 300 mg U·kg⁻¹ dw soil. However, germination apparently is not as sensitive to U toxicity as plant growth [101]. In plants grown in Hoagland's nutrient solution, Weinberger and Murthy [101] found that tomato seedlings are more sensitive to U than other plant species tested, showing a 24% reduction in growth at 4.24 mg U·L⁻¹ and 44% at 42.4 mg U·L⁻¹. Greater sensitivity of tomatoes to U in their study [101] is probably due to increased bioavailability of the U in the absence of soil. Considerable differences can be expected in the toxicity of U between studies using natural soils and nutrient solutions; the latter have little relevance to natural systems.

Addition of uranyl nitrate solution to soil suppressed the growth of spring wheat by 50% at 50 mg·kg⁻¹ soil and by 91% times at 100 mg·kg⁻¹ soil, but had no effect at 10 mg·kg⁻¹ soil. Addition of phosphorus fertilizers to soil reduces the effect of U on plant growth, probably because insoluble uranyl phosphate complexes are formed at high phosphorus (>75 µg·L⁻¹) concentrations rendering U unavailable for plant uptake [14]. In contrast, a positive response is seen in some studies when uranyl nitrate is used due to the plants responding to nitrate.

In a 22-day bioassay with northern wheatgrass (*Elymus lanceolatus*) a significant decrease in shoot and root length and in total plant biomass was observed at 3,000 mg U·kg⁻¹ soil compared to controls in neutral pH loam and for root length at 3,000 mg U·kg⁻¹ compared to controls in a fine sand with pH of 6.2 [99]. Sheppard et al. [100] reported an overall EC₂₅ in the range of 300–500 mg U·kg⁻¹ dry soil for terrestrial plants and derived a value of 250 mg U·kg⁻¹ dry soil as the Estimated-No-Effect-Concentration (ENEC).

Soil Invertebrates

The toxicity of U to soil invertebrates also has been little investigated (Table 3). In garden soil, survival of

earthworms (*Lumbricus terrestris*) to 75 days was not affected by U concentrations of 3–1,000 mg·kg⁻¹ dw soil [99]. However, the study only investigated the effects of U on mature earthworms. In another study, Environment Canada's bioassay protocol was followed for three different soil organisms: an acute 14-day survival and a 56-day chronic reproduction bioassay with the worm *Eisenia andrei*; a 28-day reproduction test with the collembolan, *Folsomia candida*; and 35-day reproduction test with mature adults of the collembolan *Onychiurus folsomi* [100]. No effect was seen on earthworm survival at 1,000 mg U·kg⁻¹ soil compared to controls for the three soils, and for reproduction in two of the three soils. An EC₂₀ of 350 mg U·kg⁻¹ soil for survival of *F. candida* was observed in limed fine-sand soil with pH 6.2. EC₂₀s of 710 to > 3,000 mg U·kg⁻¹ were recorded in the other two soils. Results for *O. folsomi* were similar with EC₂₀s from 390 to >1,000 mg U·kg⁻¹. *O. folsomi* a root-eating springtail was much more sensitive to U in the limed fine-sand soil with survival and reproduction EC₂₀s as low as 92 mg U·kg⁻¹.

Sheppard et al. [100] proposed an ENEC of 100 mg U·kg⁻¹ dry soil for the protection of soil invertebrates. This value is below the Province of Ontario [41] guideline of 300 mg U·kg⁻¹ dw soil for soil and higher than the maximum background U soil concentration [41, 102]. CCME [10] derived a soil uranium quality guideline of 23 mg·kg⁻¹ for residential, parkland, and agricultural land use for protection of both human and environmental health.

Uranium Accumulation and Toxicity in Mammals

The uptake and deposition of U in the body is dependent on the mode of entry, inhalation or ingestion, and the form of U. Early studies demonstrated that U compounds could be grouped into those that are slightly toxic (UO₂, U₃O₈, and UF₄), moderately toxic (UO₂, NaU₂O₂, (NH₄)₂U₂O₇, and high-grade ore), and highly toxic (UO₂F₂, UCl₂, UO₄). In toxicity studies, it is a common practice to inject the U intravenously to follow the fate of U once it has entered the bloodstream following ingestion and uptake by the gastrointestinal tract.

Inhalation to Lungs Inhalation is not generally considered a major pathway for uptake of U by wildlife.

However, resuspension of respirable particles in the upper few millimeters of soil may be important in the contamination of small mammals at highly contaminated sites [56]. For humans, inhalation of natural U is only important in industrial settings; however, inhalation of dust is considered the major pathway for depleted U exposure both in combat and noncombat situations [1].

Absorption of U through the lungs is related to U solubility in water and its ability to form stable bicarbonate complexes. Insoluble, UO_3 , UO_4 , UF_4 , or ammonium or sodium diuranate are deposited primarily in the lungs with little absorption into the blood. In contrast, dusts of UF_6 , UO_2F_2 , $\text{UO}_2(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$, and UCl_4 are absorbed from the lungs and circulate in the blood to bone and kidney [103]. Their effect is primarily on the kidney as described for ingestion below.

In a long-term study of up to 6.5 years duration in which dogs and monkeys were exposed for 1–5 years to natural UO_2 aerosol, Leach et al. [104] found that the lungs and the tracheobronchial lymph node were the principal sites of U retention and the only tissues that exhibited pathological effects. The kidney showed no evidence of U toxicity. At the end of 5 years of inhalation exposure, U concentrations in the lungs of dogs were in the $2,000 \mu\text{g}\cdot\text{g}^{-1}$ range ($\sim 22 \text{ mSv/week}$) and those of the monkeys were $4,000 \mu\text{g}\cdot\text{g}^{-1}$ range ($\sim 44 \text{ mSv/week}$); values well above the maximum permissible organ concentration of $25 \mu\text{g}\cdot\text{g}^{-1}$ for natural uranium [105]. They observed dose-dependent pulmonary and tracheobronchial lymph node fibrosis in the lungs, consistent with radiation effects. Effects were more marked in the monkey than in dogs. Biological half-times for U in the dog and monkey lungs ranged from 420 to 660 days. For the rat, the biological half-time in the lung averaged about 300 days.

Uptake Through the Digestive Tract The amount of U that resides in an organ at steady state is proportional to the amount of ingested U absorbed from the gut, the fraction of absorbed U deposited in the organ, and the biological half-time of U in that organ [106]. For wildlife almost all the daily U intake is from food, water, and ingestion of soil/sediment. Inhalation and dermal exposure may also be important pathways in some instances [107]. For humans, the U content of most

food is in the range of $0.01\text{--}0.1 \mu\text{g}\cdot\text{g}^{-1}$, but because of poor absorption in the gastric tract, the average daily intake of U is only about $1.0 \mu\text{g}\cdot\text{day}^{-1}$ [108]. On average, food contributes about 15% of ingested U and drinking water contributes about 85% [108].

Ingestion is probably the major exposure route to U for wildlife. A key consideration in determining the importance of ingestion is the assimilation rate from the gut. This value determines how much material is ultimately transported to the blood for deposition in critical organs and is the greatest source of uncertainty in dose estimates for the ingestion of long-lived alpha-emitting radionuclides [109]. Uptake of U following ingestion has been reviewed by Wrenn et al. [110, 111], Durbin and Wrenn [103], Yuile [107], the International Commission on Radiological Protection (ICRP) [112], and Leggett and Harrison [113] for the development of models to estimate allowable daily intakes in humans. These reviews provide data on laboratory studies with mammals and are applicable to wild species.

The uptake rate of U from the gut is generally low, ranging from $<1\%$ of the diet to 3–4% [110, 113] and can be quite low, for example, about 0.06% for both rats and rabbits with values ranging up to about 2.8% [60, 114–117]. Factors affecting this rate are the feeding status of the organism and the species [118]. The ICRP [112] assumed for adult humans, a conservative absorption rate of 5% of diet for soluble U and 0.2% for insoluble U. Slightly higher rates of uptake (up to 5%) are reported for water than food [113]. In chronic clinical studies with humans, Hursh et al. [119] reported that only about 0.5–5% of an orally administered acute dose of 11 mg U as uranium nitrate hexahydrate was absorbed. However, in another study based on daily intake rates and daily U excretion rates, absorption of 12–31% was calculated [120]. The value of 5% is considered conservative for humans. Therefore, the uptake of U by ingestion is low.

Elimination Through Feces Most of the U intake through ingestion remains in the digestive tract and is lost in the feces.

Elimination Through Kidney Absorbed U is eliminated from the body primarily by kidney excretion. More than 90% of the U absorbed into the blood is excreted into the urine and less than 1% is eliminated

by feces. Daily urinary excretion is about 0.15–0.38 μg for humans. There is a rapid initial phase of excretion in which 70–80% of the U is lost in the first 24 h following administration of the dose. The remaining 20–30% has a biological half-time of 6 months to 1 year.

The chemical toxicity of U to mammals has been reviewed by Durbin and Wrenn [103], Leggett [121], and Ribera et al. [2]. A selection of more sensitive uranium toxicity findings for mammals is presented in Table 4. The major site of U burden in mammals is the skeleton, followed by the kidney. In the skeleton, U replaces calcium in the surface of bone crystals. Effects observed in bone (i.e., osteosarcoma) are attributed to radiotoxicity and not chemical toxicity [2]. Although U accumulates in bone and kidney [103, 107, 121], chemical effects are primarily in the kidney. Literature values for the fraction of U deposited in the bone range from 20% to 32% with retention half-times of 40–180 days [114] and in the kidney range from 7.2% to 45% with retention times of 5–9 days [117]. The removal of U in the urine follows a power function $Y_u = 0.015 t^{-1.5}$ where t is in hours. Uranium deposited in the peripheral part of the renal cortex and in bones is eliminated very slowly.

Kidney Toxicity The kidney is the critical organ in terms of U chemical toxicity. This is despite the fact that much of the U in the kidney is excreted rapidly. Renal damage occurs when U reacts chemically with the protein of columnar cells lining the tubular epithelium, leading to cellular injury and necrosis [131]. In the blood, uranyl ion circulates as a relatively inert bicarbonate-uranyl complex. When this is filtered in the kidney tubules, the uranyl ions dissociate from the carbonates in the acidic environment of the kidney. The uranyl ion is concentrated in the tubular lumen where the ions may cause damage to kidney structure. If the damage is severe enough, kidney failure may occur. Necrosis of kidney tubules results in albuminuria, elevated blood urea nitrogen, and a loss of weight. The proximal convoluted tubule of the nephron is the site most affected. Uranium toxicity occurs through the disruption of kidney function by binding to the membranes of renal tubular cells, restricting the reabsorption of glucose, sodium, amino acids, protein, water and other substances, and causing cell death by suppression of cell respiration [121]. Uranyl ion also competes with ATP for a binding site on the Ca^{+} and Mg^{2+} adenosine triphosphatase (ATPase) and this may be a factor in renal toxicity [114]. Damage to the kidney

Uranium in the Environment: Behavior and Toxicity. Table 4 Selected more sensitive U toxicity findings for mammals ($\text{mg U} \cdot \text{kg}^{-1} \text{ body weight} \cdot \text{day}^{-1}$)

Taxa	Endpoint	Effect concentration	Reference
Male rat	LOEC for effects on testes	0.03	Malenchenko et al. (1978 cited in [122])
New Zealand white rabbits	LOEC for change in kidney histopathology	0.05	Gilman et al. [123, 124]
Sprague-Dawley rats	LOEC for renal damage	0.06	Gilman et al. [125]
New Zealand white rabbits	LOEC for renal damage	1.36	Gilman et al. [124]
Swiss mice	LOEC for growth of pups from exposed mothers	2.8	Paternain et al. [126]
Mice	LOEC for maternal weight gain and developmental effects	< 5	Domingo et al. [127]
Rabbit	LOEC for weight loss and limited renal damage	3	Maynard and Hodge [128]
Swiss mice	LOEC for pregnancy rate (untreated mates?)	5.6	Llobet et al. [108]
Dog	Histological effects in kidneys	9	Maynard et al. [129]
Mice	LOEC for litter size	< 50	Domingo et al. [130]

becomes evident as the cells of the proximal tubules stop functioning and die. The tubular epithelium undergoes massive cellular necrosis resulting in cell rupture and discharge of cellular contents into the urine. Cellular necrosis appears to be brought about by alternations in the transport of organic compounds and ions across tubule cell membranes along with changes in intercellular protein binding. Severe impairment may result in death. Sublethal damage is usually repairable.

Commonly used serum parameters of renal function such as creatine and urea concentration only become abnormal when a significant impairment of organic function has been sustained. Injury to the glomerulus typically increases glomerulus permeability, permitting the passage of high molecular weight proteins the size of albumin (69,000 Da) or larger to pass into the urinary filtrate. Injury to the renal proximal tubules prevents the usual reabsorption of low molecular weight proteins from the filtrate, permitting these to appear in increased concentrations in the urine, a condition called tubular proteinuria. Low molecular weight proteins are largely taken up by pinocytosis in the proximal renal tubules and then degraded to the amino acid level in the lysosomes. The proximal reabsorption of small proteins is possibly selective and seems to be a vulnerable process [114].

Toxicity of U compounds is governed by their solubility. The soluble hexavalent U compounds have LD₅₀ values after injection in the range of 0.1–1 mg U·kg⁻¹ bw in different animal species, whereas after oral application the acute toxicities are a few hundred mg·kg⁻¹ bw. Less soluble tetravalent U compounds are generally nontoxic at concentrations up to 20% of the diet.

For U, lethality is a common endpoint used for acute studies, while kidney function is the predominant endpoint for chronic studies. Acute exposures have demonstrated that U has low toxicity to most species, often requiring concentrations greater than 2% in the diet to elicit an acute response. Rabbits are a sensitive species, with LD₅₀ values below 0.5% of the diet. An LD₅₀ value of 23 mg U·kg⁻¹ bw·d⁻¹ has been reported for the rabbit in a 30-day feeding study with soluble U [103].

Animal studies performed at the Universities of Rochester and Chicago in the 1940 s and 1950 showed

that the dog is more susceptible to U⁶⁺ effects than the rodent; UO₂F₂ is the most toxic uranyl compound in acute lethality studies in rats, two times more toxic than uranyl nitrate and in long-term feeding studies was four to six times more toxic than uranyl nitrate; and U-induced renal injury (catalauria in 2–4 days) occurs in the rabbit with a dose as low as 10 µg·kg⁻¹ of U⁶⁺. Uranyl fluoride metabolism agrees generally with that of uranyl nitrate, uranyl citrate, and uranium trioxide. Renal tubular dysfunction (glucosuria) occurred in response to a 10 µg·kg⁻¹ absorbed (injected) dose. Retention half-times in the dog were 79.5 days for the kidney and 883 days for bone [132].

Environment Canada [8] in their review of U toxicity to mammals reported LD₅₀s for U toxicity administered orally of 12 mg·kg⁻¹ for dogs and 238 mg·kg⁻¹ for cats. The LD₅₀ of uranyl nitrate to rats was 135–204 mg·kg⁻¹ over 24 h and 1–2 mg·kg⁻¹ over 1–21 days. Chronic intake of U in concentrations of 0.6–1 mg·L⁻¹ causes essentially no accumulation of U in kidney or bone. The minimum oral dose given for 1 year that produced renal damage in the dog was 25 mg·kg⁻¹·day⁻¹ for uranyl fluoride; a value considerably higher than the tolerable daily intake of U of 0.6 µg·kg⁻¹ body weight from drinking water [133, 134].

Moss [114] reported that renal tissue U concentrations in the range of 1–3 mg U·kg⁻¹ lead to renal tubular changes, which are rapidly followed by adaptive and regenerative changes. Histological changes occur in rats at renal burdens of 0.7–1.3 mg·kg⁻¹. The threshold for toxic damage is as low as 0.1 mg U·kg⁻¹ bw [107, 135].

Since the kidney is the most sensitive organ to U toxicity, the threshold effect level for U in the kidney is an appropriate benchmark for the assessment of U toxicity. The ICRP [136] adopted a maximum permissible concentration of 3 mg·kg⁻¹ ww in kidney for the protection of humans, but reduced the value to 1 mg·kg⁻¹ as indicators of kidney function became more sensitive [26, 110]. SuLu and Fu-Yao Zhao [137] recommended that the kidney U burden should not be allowed to exceed 0.26 mg·kg⁻¹ in kidney of humans based on a tenfold safety factor and mild impairment of kidney function at 2.6 mg·kg⁻¹. This is in agreement with Morrow et al. [132], who found a renal concentration of 0.3 mg·kg⁻¹ kidney was close to a threshold concentration for renal injury. Damage

to kidney becomes evident in virtually all species at levels of about $0.5 \text{ mg}\cdot\text{kg}^{-1} \text{ ww}$ [26, 118]. Although the kidney has substantial reserve capacity, and may function with the loss of a proportion of the nephrons, the loss of reserve capacity is considered an adverse effect.

Based on long-term feeding studies with dogs, a No-Observed-Effect Level (NOEL) for U of $1 \text{ mg}\cdot\text{kg}^{-1}\cdot\text{day}^{-1}$ was estimated [26]. Gilman et al. [123–125] studied the effect of U on kidney damage in the rabbit and the rat. The Lowest-Observed-Effect Level (LOEL) in rabbits exposed to U as uranyl nitrate hexahydrate in their drinking water over a 91-day period was $0.05 \text{ mg U}\cdot\text{kg}^{-1} \text{ bw}\cdot\text{day}^{-1}$ ($0.04 \text{ mg U}\cdot\text{kg}^{-1} \text{ ww kidney}$) in males and $0.49 \text{ mg U}\cdot\text{kg}^{-1} \text{ bw}\cdot\text{day}^{-1}$ ($0.02 \text{ mg U}\cdot\text{kg}^{-1} \text{ ww kidney}$) in females [123]. A NOEL was not identified. In their study, four male rabbits developed *Pasteurella multocida* infection and were removed from the study. The highest doses administered, $28.7 \text{ mg U}\cdot\text{kg}^{-1}\cdot\text{day}^{-1}$ ($4.98 \text{ mg}\cdot\text{kg}^{-1} \text{ ww kidney}$) and $43.02 \text{ mg U}\cdot\text{kg}^{-1}\cdot\text{day}^{-1}$ ($1.03 \text{ mg U}\cdot\text{kg}^{-1} \text{ kidney}$) were toxic with significant effects on kidney physiology and pathology and in some cases may have been lethal [123]. Permanent kidney damage was likely in male rabbits. Female rabbits were less affected than males, although significant nuclear changes (anisokaryosis and vesiculation) were seen even at the lowest exposure ($0.49 \text{ mg U}\cdot\text{kg}^{-1}\cdot\text{day}^{-1}$). Sensitivity of male rabbits was greater in this study than in a subsequent study with specified pathogen-free male rabbits [125]. This caused Gilman et al. [125] to speculate that subclinically expressed disease stress may have caused the male rabbits to be more susceptible to the effects of U in the initial study.

In a similar study using male and female Sprague-Dawley rats, a LOEL of $0.06 \text{ mg U}\cdot\text{kg}^{-1} \text{ bw}\cdot\text{day}^{-1}$ for males and $0.09 \text{ mg U}\cdot\text{kg}^{-1} \text{ bw}\cdot\text{day}^{-1}$ for females was reported [124]. At the lowest exposure level ($0.09 \text{ mg U}\cdot\text{kg}^{-1} \text{ bw}\cdot\text{day}^{-1}$), significant changes were seen in renal tubules of males, which may result in permanent injury to basement membranes with loss of nephrons and reduced renal function [124]. In females, the most important changes were sclerosis of glomerular capsules and reticulin sclerosis of tubular basement membranes and interstitial scarring, which are non-repairable and occurred at the lowest exposure level. The authors suggested that sustained exposure would

likely have increased the number of damaged glomeruli and ultimately impaired renal function.

Effects on Offspring Domingo et al. [127] observed an increased incidence of external malformations, specifically exencephaly, cleft palate, micrognathia, and short or curled tails in response to maternally toxic doses to mice. The incidence of hematomas, especially in facial or dorsal areas, also increased. No internal soft tissue malformations were observed. Undeveloped renal papillae (a developmental alteration) occurred in fetuses from the 5 and $25 \text{ mg}\cdot\text{kg}^{-1}\cdot\text{day}^{-1}$ groups.

In another study, Paternain et al. [126] looked at the effects of uranyl acetate dehydrate on the reproduction, gestation, and postnatal survival of mice. They found that the number of dead offspring increased with oral dose. The total number of dead per total number of offspring produced per treatment over the 21-day study period expressed as a percentage was 8.2% at $0 \text{ mg}\cdot\text{kg}^{-1} \text{ bw}\cdot\text{day}^{-1}$, 23% at $5 \text{ mg}\cdot\text{kg}^{-1} \text{ bw}\cdot\text{day}^{-1}$, 26.5% at $10 \text{ mg}\cdot\text{kg}^{-1} \text{ bw}\cdot\text{day}^{-1}$, and 42.3% at $25 \text{ mg}\cdot\text{kg}^{-1} \text{ bw}\cdot\text{day}^{-1}$. The mortality of offspring between birth and day 21 (viability index) was significantly greater (Chi-squared test) in each of the U-treated groups than the controls [6]. Growth, measured as body weight over the 21-day period, was also significantly reduced in each of the U-treated groups compared with the control group. The number of dead young per litter was significantly greater than controls in the 10 and $25 \text{ mg U}\cdot\text{kg}^{-1} \text{ bw}\cdot\text{day}^{-1}$ groups at day 0.

On the basis of dead offspring per treatment over the 21-day period of lactation, mortality was three times greater at $5 \text{ mg}\cdot\text{kg}^{-1} \text{ bw}\cdot\text{day}^{-1}$ ($2.8 \text{ mg U}\cdot\text{kg}^{-1} \text{ bw}\cdot\text{day}^{-1}$) than that for controls. The value of $2.8 \text{ mg U}\cdot\text{kg}^{-1} \text{ bw}\cdot\text{day}^{-1}$ was considered a critical toxicity value (CTV) for effects on reproduction of the mouse by EC and HC [6], who used an application factor of three to derive the Estimated-No-Effect Value (ENEV) of $0.93 \text{ mg U}\cdot\text{kg}^{-1} \text{ bw}\cdot\text{day}^{-1}$ for effects on mortality of offspring in mice. The three-fourth power function for body weight [138] was used to derive species-specific ENEVs for the wildlife [6].

It can be concluded that U is a highly toxic element when soluble salts are administered peritoneally, but toxicity is lower when ingested due to low gastrointestinal absorption. Therefore, U is generally not a chemotoxic threat as an environmental contaminant

unless concentrations are elevated. Benchmark values of $0.5 \text{ mg} \cdot \text{kg}^{-1}$ ww kidney and about $1.0 \text{ mg U} \cdot \text{kg}^{-1} \text{ bw} \cdot \text{day}^{-1}$ should be protective of kidney function in most mammals.

Waterfowl

Reviews of U uptake indicate that birds may retain greater amounts of U than mammals at the same levels of exposure [139]. Sample et al. [138] reported a NOEL value of $16 \text{ mg U} \cdot \text{kg}^{-1} \text{ bw} \cdot \text{day}^{-1}$ for U toxicity to birds. This value was based on a toxicity study by Haseltine and Sileo [140] on the black duck using depleted U. However, the use of this study to derive a NOEL for birds is questionable, because depleted U is a metallic form of U which has very low bioavailability in the gut.

Growing chicks are relatively tolerant of U with a 7-day LD_{50} of $235 \text{ mg} \cdot \text{kg}^{-1}$ body weight, which is considerably higher than 24-h LD_{50} of 135–305 $\text{mg} \cdot \text{kg}^{-1}$ bw for rats [141]. Hepatotoxicity is a predominant feature of U toxicity in chickens.

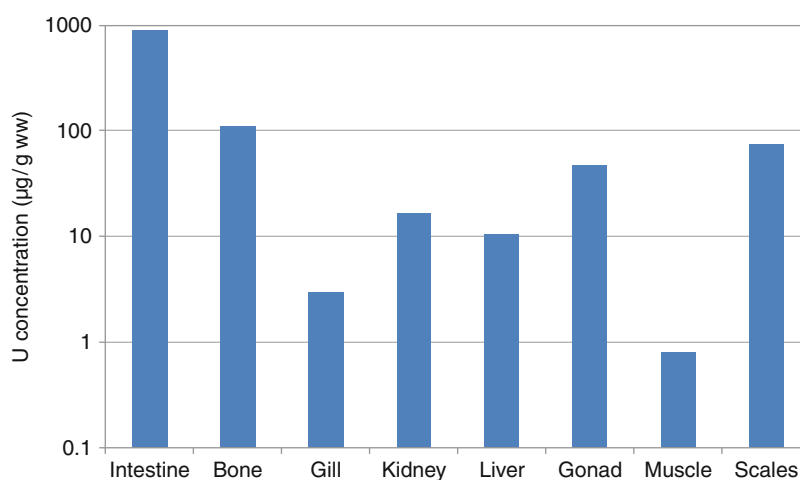
Fish

Absorption of U by fish takes place in the gut, epidermis, and gills. Uptake of U via the diet is considered to be the primary route of U exposure for fish [68, 142–144], although others consider uptake to be primarily via the gill as is common for metals [145]. In laboratory

studies, Simon and Garnier-Laplace [146] found that trophic transfer through the diet was minimal as is the case for many terrestrial species. Uranium accumulates in intestine, bone, scales, gonad, kidney, liver, and gill when U is taken up through the gut from the diet (Fig. 2).

The renal tubular excretion system in fish is similar to other vertebrates [8]; therefore, the toxicity of U to fish is expected to be similar to that of mammals. For example, lake whitefish (*Coregonus clupeaformis*) fed a commercial diet contaminated with U developed renal lesions in proximal tubules similar to those reported for mammals [143]. Focal hepatocyte necrosis of the liver and alterations of the bile ductile epithelium were also observed.

Acute and chronic toxicity data for aquatic biota are summarized in Table 5. In several studies, toxicity was greater in soft water than in hard water. For example, Tarzwell and Henderson [148] reported LC_{50} s for fat-head minnow (*Pimephales promelas*) of $2.8 \text{ mg} \cdot \text{L}^{-1}$ in soft water ($20 \text{ mg} \cdot \text{L}^{-1}$ hardness) and $135 \text{ mg} \cdot \text{L}^{-1}$ in hard water ($400 \text{ mg} \cdot \text{L}^{-1}$ hardness) for uranyl sulfate. Parkhurst et al. [149] reported a sixfold increase in hardness from 35 to $208 \text{ mg} \cdot \text{L}^{-1}$ decreased the 96-h LC_{50} of U to brook trout (*Salvelinus fontinalis*) by a factor of four. In soft water, U LC_{50} values ranged from 0.73 to $3.46 \text{ mg} \cdot \text{L}^{-1}$ for six tropical species of fish [145], while the threshold toxicity for growth in a larval



Uranium in the Environment: Behavior and Toxicity. Figure 2

Uranium concentration in tissues of white fish exposed to contaminated diet for 100 days (Based on Cooley and Klaverkamp [144])

Uranium in the Environment: Behavior and Toxicity. Table 5 Uranium toxicity ($\mu\text{g}\cdot\text{L}^{-1}$) to freshwater aquatic organisms with emphasis on hardness as a modifying parameter. *H* hardness as $\text{mg}\cdot\text{L}^{-1}$ CaCO_3 , *A* alkalinity as $\text{mg}\cdot\text{L}^{-1}$ CaCO_3

Taxa	Endpoint	Aqueous parameters	Effect concentration ($\mu\text{g}\cdot\text{L}^{-1}$)	Reference and true hardness assessed yes or no ^a
Fish				
Fathead minnow (<i>Pimephales promelas</i>)	96-h LC ₅₀	H23, A12, pH 6.3–7.0	2,000	Vizon SciTec Inc [147] Yes, no effect
	96-h LC ₅₀	H72, A14, pH 6.5–6.8	2,000	
	96-h LC ₅₀	H131, A10, pH 6.4–6.8	2,100	
	96-h LC ₅₀	H244, A10, pH 6.4–6.8	1,800	
	7-day LC ₅₀	H23, A12, pH 6.3–7.0	1,600	
	7-day LC ₅₀	H72, A14, pH 6.5–6.8	2,100	
	7-day LC ₅₀	H131, A10, pH 6.4–6.8	2,000	
	7-day LC ₅₀	H244, A10, pH 6.4–6.8	1,500	
	7-day LOEC		1,300–2,000	
	96-h LC ₅₀	H20, A18, pH 7.4	1,572	
Fathead minnow (<i>Pimephales promelas</i>)	96-h LC ₅₀	H400, A360, pH 8.2	75,778	Tarzwel and Henderson [148] No
Brook trout (<i>Salvelinus fontinalis</i>)	48-h LC ₅₀	H184, A146, pH 7.4	59,000	Parkhurst et al. [149] No
	96-h LC ₅₀	H32, A12, pH 6.7	5,500	
	96-h LC ₅₀	H210, A54, pH 7.5	23,000	
	60-day embryo larval NOEC	H201, A189, pH 8.0	> 9,080	
	96-h LC ₅₀	H31, A26, pH 6.8–7.0	8,000	
Rainbow trout (<i>Oncorhynchus mykiss</i>)	96-h LC ₅₀	H20, A20, pH 6.2–7.0	4,200	Vizon SciTec Inc [147] Yes, no effect
	96-h LC ₅₀	H68, A12, pH 6.6–7.0	3,900	
	96-h LC ₅₀	H126, A11, pH 6.5–6.9	4,000	
	96-h LC ₅₀	H243, A12, pH 6.5–7	3,800	
	30-day EC ₅₀ , NOEL/LOEL embryo/alevin	H6/A7, pH 6.3–7.2	460 <280 280	
Rainbow trout (<i>Oncorhynchus mykiss</i>)	31-day EC ₅₀ embryo/alevin NOEL/LOEL	H61/A6, pH 6.3–7.2	640 310 610	Vizon SciTec Inc [147] Yes, minor effect

Uranium in the Environment: Behavior and Toxicity. Table 5 (Continued)

Taxa	Endpoint	Aqueous parameters	Effect concentration ($\mu\text{g}\cdot\text{L}^{-1}$)	Reference and true hardness assessed yes or no ^a
Rainbow trout (<i>Oncorhynchus mykiss</i>)	96-h LC ₅₀	H31, A26, pH 6.8–7.0	6,200	Davies [150] No
Lake trout (<i>Salvelinus namaycush</i>)	141-day NOEC 141-day LOEC embryo/alevin/fry	H74–88, A69–77, pH 7.9–8.1	6,050 29,780	Liber et al. [151] No
Bluegill (<i>Lepomis macrochirus</i>)	96-h LC ₅₀	H2.5–3.2, A < 0.1–3.8, pH 5.1–5.6	1,670	Trapp [152] No
Zebra fish (<i>Brachydanio rerio</i>)	96-h LC ₅₀	H178, pH 7.9	3,050	Labrot et al. [153] No
Colorado squawfish (<i>Ptychocheilus lucius</i>)	96-h LC ₅₀ Swim-up, fry, small, and larger juvenile	H196, A107, pH 7.8	46,000 all stages	Hamilton [154] No
Razorback sucker (<i>Xyrauchen texanus</i>)	96-h LC ₅₀ Swim-up, fry, small, and larger juvenile	H196, A107, pH 7.8	46,000	Hamilton [154] No
Bonytail (<i>Gila elegans</i>)	96-h LC ₅₀ Swim-up, fry, small, and larger juvenile	H196, A107, pH 7.8	46,000	Hamilton [154] No
Flannelmouth sucker (<i>Catostomus latipinnis</i>)	96-h LC ₅₀	H144, A103, pH 7.9	43,500	Hamilton and Buhl [155] No
Chequered rainbowfish (<i>Melanotaenia splendida inornata</i>)	96-h LC ₅₀ 7-day old 96-h LC ₅₀ 14-day old 96-h LC ₅₀ 90-day old	H4.6, A3.3, pH 6.6 H4, A3.2, pH 6.6 H4.6, A3.3, pH 6.6	2,660 1,390 3,460	Bywater et al. [145] No
Chequered rainbowfish (<i>Melanotaenia splendida inornata</i>)	96-h LC ₅₀ 14-day old 7-day LC ₅₀ 31-day old	H4.1, A1.8, pH 6.3 H4.1, A1.8, pH 6.3	1,390 1,570	Holdway [156] No
Black-banded rainbowfish (<i>Melanotaenia nigrans</i>)	96-h LC ₅₀ 7-day old 96-h LC ₅₀ 90-day old	H4.7, A3.3, pH 6.6 H4.7, A3.3, pH 6.6	1,700 1,900	Bywater et al. [145] No
Northern purple-spotted gudgeon (<i>Mogurnda mogurnda</i>)	96-h LC ₅₀ 6-day old 96-h LC ₅₀ 7-day old 96-h LC ₅₀ 40-day old 96-h LC ₅₀ 70-day old 96-h LC ₅₀ 90-day old	H4, A3.2, pH 6.6 H4.6, A3.3, pH 6.6 H4, A3.2, pH 6.6 H4, A3.2, pH 6.6 H4.6, A3.3, pH 6.6	1,570 1,110 329 329 1,460	Bywater et al. [145] No

Northern purple-spotted gudgeon (<i>Mogurnda mogurnda</i>)	96-h LC ₅₀ 6-day old	H4, A3.2, pH 6.6	1,570	Holdway [156] No
	96-h LC ₅₀ 40-day old	H4, A3.2, pH 6.6	3,290	
	96-h LC ₅₀ 70-day old	H4, A3.2, pH 6.6	3,290	
	7-day LC ₅₀ 1-day old	H4, A1.8, pH 6.3	1,590	
	7-day LC ₅₀ 1-day old + 7-day postexposure	H4, A1.8, pH 6.3	890	
	7-day LC ₅₀ 40-day old	H4, A1.8, pH 6.3	2,690	
	7-day LC ₅₀ 40-day old + 7-day postexposure	H4, A1.8, pH 6.3	1,440	
	7-day LC ₅₀ 70-day old	H4, A1.8, pH 6.3	3,290	
	7-day LC ₅₀ 70-day old + 7-day postexposure	H4, A1.8, pH 6.3	2,700	
	14-day LC ₅₀ , 1-day old	H3.6, A3, pH 6.4	>1,790	
	14-day LC ₅₀ + 15-day postexposure	H3.6, A3, pH 6.4	1,290	
	14-day LOEC, 1-day old		1,790	
	14-day NOEC, 1-day old		880	
Reticulated perchler (<i>Ambassis macleayi</i>)	96-h LC ₅₀ juvenile	H4.6, A3.3, pH 6.6	800	Bywater et al. [145] No
Delicate blue-eyes (<i>Pseudomugil tenellus</i>)	96-h LC ₅₀ juvenile	H4.6, A3.3, pH 6.6	730	Bywater et al. [145] No
Mariana's Hardyhead	96-h LC ₅₀ juvenile	H4.6, A3.3, pH 6.6	1,220	Bywater et al. [145] No
Zooplankton				
Water flea (<i>Daphnia magna</i>)	48-h LC ₅₀	H67-73, A54-60, pH 8	5,340–7,620	Poston et al. [157] No
	48-h LC ₅₀	H126–140, A93, pH 8	30,440–44,570	
	48-h LC ₅₀ survival	H188–205, A124–133, pH 8	29,560–74,340	
	21-day LOEC reproduction	H200, pH 7.8–8	520	
Water flea (<i>Daphnia magna</i>)	48-h LC ₅₀	H91, A62, pH 7.7	5,180–9,360	Barata et al. [36] No
		H179, A126, pH 8.1	15,300–25,400	

Uranium in the Environment: Behavior and Toxicity. Table 5 (Continued)

Taxa	Endpoint	Aqueous parameters	Effect concentration ($\mu\text{g}\cdot\text{L}^{-1}$)	Reference and true hardness assessed yes or no ^a
Water flea (<i>Daphnia magna</i>)	48-h LC ₅₀	H250, A2.7, pH 7	390	Zeman et al. [158]
		H250, A34, pH 8	7,800	
	21-day LC ₁₀	H250, A2.7, pH 7	100	
	21-day EC ₅₀ reproduction	H250, A2.7, pH 7	91	
	21-day EC ₁₀ reproduction	H250, A2.7, pH 7	14	
Water flea (<i>Daphnia pulex</i>)	48-h LC ₅₀	H2.3–3.3, A < 0.1–0.6, pH 5.1–5.6	220	Trapp [152] No
Water flea (<i>Ceriodaphnia dubia</i>)	96-h LC ₅₀	H176, A126, pH 8.4	10,500	Kuhne et al. [159] No
	7-day NOEC	H190, A148, pH 8.5	1,970	
	7-day LOEC	H190, A148, pH 8.5	3,910	
Water flea (<i>Ceriodaphnia dubia</i>)	IC ₅₀ LOEC NOEC reproduction	H60, A22, pH 7.7	44 35 10	Liber [160] No
	IC ₅₀ LOEC NOEC reproduction	H120, A69, pH 8.3	> 1,147 380 127	
	IC ₅₀ LOEC NOEC reproduction	H240, A141, pH 8.6	> 4,295 4,295	
	3-brood LC ₅₀ LC ₂₅	H5, A5, pH 6.5–6.9	160 150	
Water flea (<i>Ceriodaphnia dubia</i>)	3-brood LC ₅₀ LC ₂₅	H17, A5, pH 6.6–7	140 120	Vizon Scitec Inc [147] Yes, No effect
	3-brood LC ₅₀ LC ₂₅	H124, A8, pH 6.6–7.3	100 58	
	3-brood LC ₅₀ LC ₂₅	H252, A8, pH 6.7–7.1	110 54	
	48-h LC ₅₀ NOEC MATC	H6.1, A1.1, pH 6.9–7.7	60 and 89 two tests < 1.3, < 7, 2.5 2	
<i>Diaphanosoma exisum</i>	24-h LC ₅₀	H4.6, A3.3, pH 6.6	1,000	Bywater et al. [145] No
<i>Latonopsis fasciculate</i>	24-h LC ₅₀	H4.6, A3.3, pH 6.6	410	Bywater et al. [145] No
<i>Dadaya macrops</i>	24-h LC ₅₀	H4.6, A3.3, pH 6.6	1,100	Bywater et al. [145] No
Water flea (<i>Moinodaphnia macleayi</i>)	48-h LC ₅₀	H4.6, A3.3, pH 6.6	1,290	Bywater et al. [145] No

Water flea (<i>Moinodaphnia macleayi</i>)	48-h EC ₅₀	pH 6.3–6.9	160–390	Semaan et al. [162] No
	48-h LOEC		100–370	
	48-h NOEC		100–270	
	5–6 day LOEC reproduction	pH 6.85	20– >46	
	5–6 day NOEC reproduction		8–46	
Phytoplankton				
<i>Chlorella</i> sp.	72-h EC ₅₀ LOEC NOEC	H8, A8, pH 7	56 1.6 0.7	Charles et al. [163] Yes
	72-h EC ₅₀ LOEC NOEC	H40, A8, pH 7	72 1.7 0.7	
	72-h EC ₅₀ LOEC NOEC	H100, A8, pH 7	150 4.4 2.3	
	72-h EC ₅₀ LOEC NOEC growth inhibition	H400, A8, pH 7	270 12 4.5	
<i>Chlorella</i> sp.	72-h EC ₅₀ LOEC NOEC	H4, pH 5.7	78 34 21	Franklin et al. [164]
	72-h EC ₅₀ LOEC NOEC growth inhibition	H4, pH 6.5	44 13 11	
<i>Chlorella</i> sp.	72-h IC ₅₀ NOEC LOEC growth synthetic water	H3.9, A4, pH 6.2	74 38 70	Hogan et al. [165] 4 reps
	72-h IC ₅₀ NOEC LOEC growth natural water		177 150 179	
			166 109 136	
			238 157 187	
<i>Pseudokirch neriella subcapitata</i> (previously <i>Selenastrum capricornutum</i>)	72-h IC ₅₀ IC ₂₅ NOEL LOEL	H5, A7, pH 6.8–7.8	160 27 14 29	Vizon SciTec Inc [147] Yes, no, or little effect
	72-h IC ₅₀ IC ₂₅ NOEL LOEL	H15, A7, pH 7.8–8.2	170 94 57 110	
	72-h IC ₅₀ IC ₂₅ NOEL LOEL	H64, A7, pH 7.2–7.3	100 60 56 110	
	72-h IC ₅₀ IC ₂₅ NOEL LOEL	H122, A8, pH 7.2–7.4	200 100 220 430	
	72-h IC ₅₀ IC ₂₅ NOEL LOEL	H228, A8, pH 7.2–7.3	200 150 110 210	
Macrophytes				
Duck weed (<i>Lemna minor</i>)	7-day IC ₅₀ IC ₂₅ dw	H35, A7, pH 5.8–7.4	13,100 6,400	Vizon SciTec Inc [147] Yes, effect
	7-day IC ₅₀ IC ₂₅ dw	H137, A9, pH 6.5–7.3	35,500 3,300	
	7-day IC ₅₀ IC ₂₅ frond #	H35, A7, pH 5.8–7.4	7,400 4,700	
	7-day EC ₅₀ EC ₂₅ frond #	H137, A9, pH 6.5–7.3	16,400 12,300	

Uranium in the Environment: Behavior and Toxicity. Table 5 (Continued)

Taxa	Endpoint	Aqueous parameters	Effect concentration ($\mu\text{g}\cdot\text{L}^{-1}$)	Reference and true hardness assessed yes or no ^a
Duck weed (<i>Lemna aequinoctialis</i>)	IC ₅₀		> 2,850	Van Dam et al. [166]
	IC ₁₀		250	
Benthic invertebrates				
Freshwater snail (<i>Amerianna cumingi</i>)	IC ₅₀		250	Van Dam et al. [166]
	IC ₁₀		22	
Amphipod (<i>Hyalella azteca</i>)	14-day LC ₅₀ EC ₂₅ NOEC LOEC	H17, A8, pH 6.4–7.0	17 5.3 5.7 11	Vizon SciTec Inc [147] yes, effect
	14-day LC ₅₀ EC ₂₅ NOEC LOEC	H61, A8, pH 6.4–6.9	140 100 70 170	
	14-day LC ₅₀ EC ₂₅ NOEC LOEC	H123, A9, pH 6.5–7.1	200 130 63 130	
	14-day LC ₅₀ EC ₂₅ NOEC LOEC	H238, A10, pH 6.5–7.0	340 130 94 190	
Amphipod (<i>Hyalella azteca</i>)	14-day LC ₅₀ Depleted U – low bioavailability	H157, A137, pH 7.9	1,520	Kuhne et al. [159] No
Amphipod (<i>Hyalella azteca</i>)	7-day LC ₅₀	H18, A14, pH 7.4–8.3	21	Borgmann et al. [167] No
	7-day LC ₅₀	H124, A84, pH 8.2–8.5	1,651 nominal value	
Amphipod (<i>Hyalella azteca</i>)	Adults		Adults	Alves et al. [168]
	28-day LC ₅₀	H12, A8.4, pH 6.9	42	
	28-day LC ₅₀	H12, A8.4, pH 7.2	15	
	28-day LC ₅₀	H120, A8.4, pH 7.2	30	
	28-day LC ₅₀	H60, A42, pH 7.6	48	
	28-day LC ₅₀	H12, A84, pH 8.0	94	
	28-day LC ₅₀	H120, A84, pH 7.9	1,042	
	28-day LC ₅₀	H120, A84, pH 8	2,066	
	Juveniles		Juveniles	
	28-day LC ₅₀	H12, A8.4, pH 6.9	14.5	
Amphipod (<i>Hyalella azteca</i>)	28-day LC ₅₀	H12, A8.4, pH 7.2	47	
	28-day LC ₅₀	H120, A8.4, pH 7.2	44	

	28-day LC ₅₀	H60, A42, pH 7.6	198	Burnett and Liber [169] No
	28-day LC ₅₀	H12, A84, pH 8.0	132	
	28-day LC ₅₀	H120, A84, pH 7.9	664	
	28-day LC ₅₀	H120, A84, pH 8	714	
	Growth		Growth	
	28-day EC ₂₅	H12, A8.4, pH 6.9	11	
	28-day EC ₂₅	H12, A8.4, pH 7.2	21	
	28-day EC ₂₅	H120, A8.4, pH 7.2	28	
	28-day EC ₂₅	H60, A42, pH 7.6	59	
	28-day EC ₂₅	H12, A84, pH 8.0	45	
	28-day EC ₂₅	H120, A84, pH 7.9	595	
Midge (<i>Chironomus tentans</i>)	28-day EC ₂₅	H120, A84, pH 8	262	Dias et al. [170]
	10-day EC ₅₀	H125, A84, pH 7.2	10,200	
Chironomid larvae (<i>Chironomus riparius</i>)	MATC		800	Markich et al. [29]
	10-day ~ LC ₇₂ 6.07 µg U·g ⁻¹ sediment dw	H48, pH 7.2	28	
	10-day ~ LC ₇₇ 11.4 µg U·g ⁻¹ sediment dw	H48, pH 7.2	51	
	10-day ~ LC ₇₇ 11.4 µg U·g ⁻¹ sediment dw	H48, pH 7.2	85	
	EC ₅₀ , NOEC, LOEC	H3.7, A2.7 pH 5	117 95 92	
	EC ₅₀ , NOEC, LOEC	H3.7, A2.7 pH 5	144 118 113	
	EC ₅₀ , NOEC, LOEC	H3.7, A2.7 pH 5	247 208 197	
	EC ₅₀ , NOEC, LOEC	H3.7, A2.7 pH 5.3	141 114 108	
	EC ₅₀ , NOEC, LOEC	H5.5, A2.7 pH 5.5	163 132 125	
	EC ₅₀ , NOEC, LOEC	H5.5, A2.7 pH 5.5	242 228 192	
Bivalve (<i>Vesunio angasi</i>)	EC ₅₀ , NOEC, LOEC	H5.5, A2.7 pH 5.5	497 417 399	
	EC ₅₀ , NOEC, LOEC	H5.8, A2.7 pH 5.5	290 224 214	

Uranium in the Environment: Behavior and Toxicity. Table 5 (Continued)

Taxa	Endpoint	Aqueous parameters	Effect concentration ($\mu\text{g}\cdot\text{L}^{-1}$)	Reference and true hardness assessed yes or no ^a
Bivalve (<i>Corbicula fluminea</i>)	EC ₅₀ , NOEC, LOEC	H6.0, A2.7 pH 5.5	634 440 416	
	EC ₅₀ , NOEC, LOEC	H6.0, A2.7 pH 5.5	824 598 558	
	EC ₅₀ , NOEC, LOEC valve movement	H6.0, A2.7 pH 5.5	1,228 958 913	
	EC ₅₀ , EC ₂₀	H6, pH 5.5, 10 min	40 19	Fournier et al. [33]
	EC ₅₀ , EC ₂₀	H6, pH 5.5, 20 min	40 19	
	EC ₅₀ , EC ₂₀	H6, pH 5.5, 30 min	40 19	
	EC ₅₀ , EC ₂₀	H6, pH 5.5, 60 min	40 19	
	EC ₅₀ , EC ₂₀	H6, pH 5.5, 120 min	28.6 <DL	
	EC ₅₀ , EC ₂₀	H6, pH 5.5, 300 min	11.9 <DL	
	EC ₅₀ , EC ₂₀	H5.9, pH 6.5, 10 min	150 90	
	EC ₅₀ , EC ₂₀	H5.9, pH 6.5, 20 min	143 79	
	EC ₅₀ , EC ₂₀	H5.9, pH 6.5, 30 min	138 52	
Hydra (<i>Hydra viridissima</i>)	EC ₅₀ , EC ₂₀	H5.9, pH 6.5, 60 min	83 4.8	Rietmuller et al. [35]
	EC ₅₀ , EC ₂₀	H5.9, pH 6.5, 120 min	38 <DL	
	EC ₅₀ , EC ₂₀	H5.9, pH 6.5, 300 min	31 <DL	
	96-h EC ₅₀ , LOEC	H6.6, A4, pH 6	114 32	
	96-h EC ₅₀ , LOEC	H165, A4, pH 6	177 90	
Hydra (<i>Hydra viridissima</i> or <i>H. vulgaris</i>)	96-h EC ₅₀ , LOEC	H165, A102, pH 6	171 42	
	96-h EC ₅₀ , LOEC	H330, A4, pH 6	219 62	
	LOEC		400–550	
				Hyne et al. [171]

^aTrue hardness – varying the hardness concentration without varying alkalinity, so alkalinity does not covary with hardness.

tropical fish was reported at $0.20 \text{ mg}\cdot\text{L}^{-1}$ [156]. In very soft water, bluegill (*Lepomis macrochirus*) is also sensitive with a 96-h LC_{50} of $1.7 \text{ mg}\cdot\text{L}^{-1}$ [152]. Hamilton [154] reported that the toxicity of U to the Colorado squawfish (*Ptychocheilus lucius*), razorback sucker (*Xtrauchen texanus*), and bonytail (*Gila elegans*) is low (LC_{50} of $46 \text{ mg}\cdot\text{L}^{-1}$) in hard water (Table 5). Hamilton [154] found that the acute toxicity of the three fishes to U was the same for the three early life stages tested: swim-up fry, young juvenile, and older juvenile. This is in contrast to most studies where early life stages are usually more sensitive to toxicants than later life stages. Hamilton [154] also found that toxicity to fish was in the order vanadium = zinc > selenite > lithium = uranium > selenate > boron. Uranium was also more toxic than lead to the zebra fish *Brachydanio rerio* [153].

Rainbow trout (*O. mykiss*) appear to be moderately tolerant of U with 96-h LC_{50} s ranging from 3.8 to $6.2 \text{ mg}\cdot\text{L}^{-1}$ [147, 150] at various levels of hardness (Table 5). However, rainbow trout embryos and alvins were sensitive to U in moderately hard water with low alkalinity in longer-term exposures. Based on survival after 31 days, the LOEC and EC_{25} were $280 \text{ }\mu\text{g}\cdot\text{L}^{-1}$ and $340 \text{ }\mu\text{g}\cdot\text{L}^{-1}$ respectively [147]. In another study, lake trout (*S. namaycush*) embryos and fry had a LOEC of $29.78 \text{ mg U}\cdot\text{L}^{-1}$ for time to hatch, hatching success, and survival in moderately hard water ($72 \text{ mg CaCO}_3\cdot\text{L}^{-1}$) after 141 days of exposure but the exposure had no effect on growth [151]. These studies demonstrate that there is a marked difference in sensitivity to U for the two trout species in water that differed primarily in the level of alkalinity.

For fathead minnow (*P. promelas*) embryos in 7-day exposure tests, NOECs ranged from 810 to $1,200 \text{ }\mu\text{g}\cdot\text{L}^{-1}$ and LOECs from $1,300$ to $2,000 \text{ }\mu\text{g}\cdot\text{L}^{-1}$ over a range of water hardness [147]. White sucker (*C. commersoni*) fry in moderately hard water ($72 \text{ mg CaCO}_3\cdot\text{L}^{-1}$) were relatively tolerant of U with a NOEC of $7.33 \text{ mg}\cdot\text{L}^{-1}$ and a LOEC of $27.86 \text{ mg}\cdot\text{L}^{-1}$ for growth [172]. In general, as the duration of the test increases, toxicity occurs at lower U concentrations for the same species and endpoints, although little difference was observed between 96-h and 7-day LC_{50} values for the fathead minnow [147].

It is sometimes suggested that in areas of elevated U concentrations, the local fauna will adapt to local conditions and become more tolerant to U. To test this

conjecture, Keklak et al. [173] compared the toxicity of U to second-generation mosquitofish from parents chronically exposed to elevated aqueous concentrations of $4.2 \text{ }\mu\text{g U}\cdot\text{L}^{-1}$ and sediment concentrations of $6,165 \text{ }\mu\text{g U}\cdot\text{g}^{-1} \text{ dw}$ with that of specimens from a reference site. After 7 days of exposure to $2.57 \text{ mg}\cdot\text{L}^{-1}$ of U as uranyl nitrate, 98% and 96% of the reference population had died in two replicate tanks compared to 25% and 57% of the offspring of fish that had been exposed for multiple generations. They concluded that the increased tolerance to U was genetically based since the test organisms were second-generation postexposure. However, this is not a substantial increase in tolerance and from data presented in their paper, it can be seen that the mortality of exposed fish was still increasing and had the experiment continued, all exposed fish likely would have died.

Zooplankton

Pickett et al. [161] found that *Ceriodaphnia dubia* was very sensitive to U with a threshold effect value of $3 \text{ }\mu\text{g}\cdot\text{L}^{-1}$ (GM of the NOEC and LOEC) for reproduction in soft water ($\sim 4 \text{ mg}\cdot\text{L}^{-1} \text{ CaCO}_3$) (Table 5). In another study, *C. dubia* had a 48-h threshold effect concentration for reproduction of $19 \pm 2 \text{ }\mu\text{g U}\cdot\text{L}^{-1}$ in 7-day chronic toxicity tests with water from a stream near McClean Lake Operations in northern Saskatchewan, adjusted to a hardness of $60 \text{ mg}\cdot\text{L}^{-1}$ as CaCO_3 [160]. When the hardness was adjusted to $120 \text{ mg}\cdot\text{L}^{-1} \text{ CaCO}_3$, toxicity was considerably lower with a threshold effect concentration for reproduction of $218 \text{ }\mu\text{g U}\cdot\text{L}^{-1}$ [160]. Vizon SciTec [147] reported LC_{25} s of $54\text{--}150 \text{ }\mu\text{g}\cdot\text{L}^{-1}$ for *C. dubia* at hardness of $5\text{--}252 \text{ mg}\cdot\text{L}^{-1}$ as CaCO_3 and low alkalinity ($5\text{--}8 \text{ mg}\cdot\text{L}^{-1}$ as CaCO_3). In contrast, Kuhne et al. [159] reported low toxicity in hard water ($182 \text{ mg}\cdot\text{L}^{-1}$ as CaCO_3) for *C. dubia* with a 96-h LC_{50} of $10.5 \text{ mg}\cdot\text{L}^{-1}$ and a LOEC of $\geq 3.91 \text{ mg}\cdot\text{L}^{-1}$ for depleted U.

Poston et al. [157] reported that the 48-h LC_{50} for *D. magna* to U decreased sixfold as hardness increased from 70 to $195 \text{ mg}\cdot\text{L}^{-1}$, whereas Barata et al. [36] reported a threefold decrease in toxicity as hardness increased from 90 to $180 \text{ mg}\cdot\text{L}^{-1}$. The LOEL was $520 \text{ }\mu\text{g}\cdot\text{L}^{-1}$ for daphnid reproduction in water with hardness of $66\text{--}73 \text{ mg}\cdot\text{L}^{-1}$ [157], whereas the 48-h LC_{50} for *Daphnia pulex* in soft water was $220 \text{ }\mu\text{g}\cdot\text{L}^{-1}$ [152].

In a more detailed study on the effect of U on *D. magna*, Zeman et al. [158] reported a 48-h LC₅₀ of 390 µg·L⁻¹ at pH 7 and 7,800 µg·L⁻¹ at pH 8. Much greater sensitivity was observed in 21-day exposures. The 21-day LC₁₀ was 100 µg·L⁻¹ and the EC₁₀ for reproduction was 14 µg·L⁻¹. Decreased feeding activity and increased oxygen consumption at 25 µg·L⁻¹ of U after 7 days exposure affected growth. Growth and reproduction decreased 50% and 65%, respectively, at 50 µg·L⁻¹ of U after 7 days and at 25 µg·L⁻¹ of U after 21 days. Therefore, *D. magna* is sensitive to low concentrations of U, but not as sensitive as *C. dubia*.

Bywater et al. [145] reported U LC₅₀ values ranging from 0.14 to 0.9 mg·L⁻¹ for four species of tropical cladocerans in water of low hardness and low alkalinity. LOECs ranged from 20 to 49 µg·L⁻¹ and NOECs from 8 to 31 µg·L⁻¹ for three different populations of the tropical cladoceran *Moinodaphnia macleayi* [162]. The population of *M. macleayi* obtained from a pond with elevated U concentrations was found to be no more tolerant than the populations from background conditions.

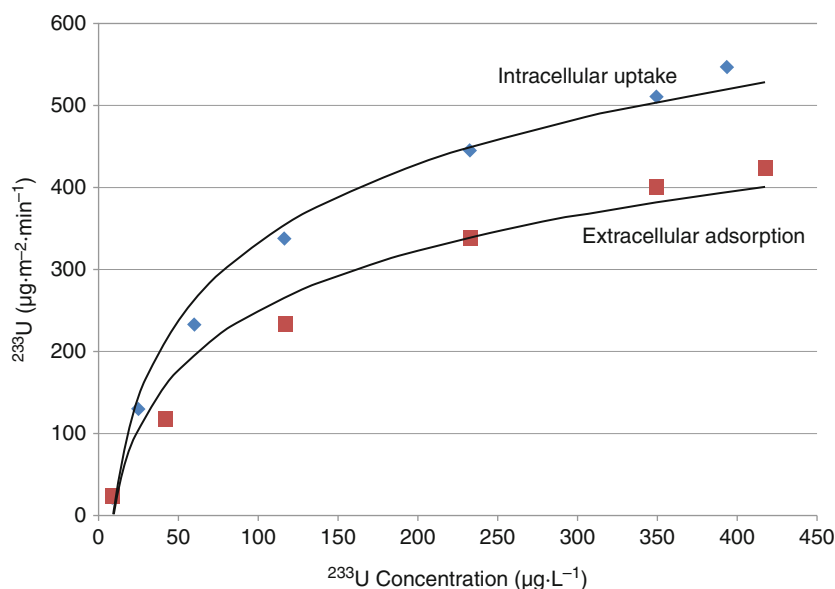
Phytoplankton

Cellular Uptake of U Uranium is believed to mimic calcium at the cell surface and is readily transported into the cell as if it were calcium. Once inside the algal cell, U is concentrated into vacuoles and may be precipitated as U(VI) phosphates [7]. Internal deposition of U may have ultimately killed the cells of *Pseudomonas fluorescens* [174]. Cellular toxicity may occur as a result of binding of U to enzymes disabling enzyme function [175]. Uranium may also exert toxicity at the cellular level by interfering with calcium homeostasis. Uptake and deposition of U in marine red algae is primarily by ion exchange or coprecipitation of the U ion with the calcium carbonate matrix and secondarily by complex formation with either protein nitrogen or other components of the organic fraction. Competition for sites on cell walls is strongly pH dependent as H⁺ ions behave like other cations. At pH below 5, H⁺ concentrations are high enough to compete with other metals in solution and strip metals from the ligands. Uranium(IV) also competes with other metals for binding sites based on molecular size and shape, and the configuration of the ligand.

Algae are very efficient at removing nutrients and metals, including U, from solution. Iron reduces the uptake of U by *Anacystis nidulans* (Cyanobacteria) [176], as do phosphate and carbonate for U uptake by *Chlorella regularis* (Chlorophyta) [177]. Adsorption of U and other metals (Cu, Ni, Mn, and Zn) reaches steady state rapidly, within a few minutes. Uranium is concentrated from water by a factor of 330 in 48 h by the alga *Ochromonas* sp. B and by a factor of 2,098 based on wet weight by the green alga *Spirogyra* sp. in a U enriched stream (12.5 µg·L⁻¹) [6]. Both living and nonliving *Chlorella regularis* accumulate U from water [178]. Figure 3 shows the uptake of U by algae after 30 min due to intracellular uptake and extracellular adsorption. Intracellular U content of cells was greatly reduced in the presence of ligands. About 75% of the intracellular U was associated with the insoluble fractions – (vacuoles, nuclei, flagella, and membranes), and 25% with the cytosol [32].

Algal cells take up U in the cation form UO₂²⁺ or UO₃OH⁺ [177]. The optimum pH for U uptake by algae is usually acidic to neutral. At pH 5.9–6.8 uranyl ions exist as hydrolysis products, which are more strongly bound to cell absorption sites. Above pH 7, uranyl carbonate complexes exist. These negatively charged and extremely soluble ionic complexes experience charge conflicts at cell absorption sites, which inhibit uptake thereby decreasing U uptake. Optimum pH range for U uptake by algae is about pH 2.5–8.5 (Table 6) and between pH 4 and 5 for lichen. Uptake optima under acidic conditions correspond to the fact that carboxyl, amino, and phosphate ligands are probably the major complexing agent on or in the cell walls.

Toxicity to Phytoplankton Franklin et al. [164] found that a tropical alga (*Chlorella* sp.) was more sensitive to U in soft water (2–4 mg·L⁻¹ as CaCO₃) at pH 6.5 than at pH 5.7 (Table 5). In their study, the LOEC was 34 µg·L⁻¹ at pH 5.7 and 13 µg·L⁻¹ at pH 6.5. At pH 5.7, U concentrations up to 34 µg·L⁻¹ had no significant effect on growth rate, but at 60 µg·L⁻¹ relatively small increases in U concentrations caused pronounced growth inhibition. At pH 6.5, *Chlorella* sp. was more sensitive to U with concentrations >8 µg·L⁻¹ resulting in pronounced growth inhibition with small increases in U concentration. Increasing the U concentration from 40 to 250 µg·L⁻¹ resulted in an



Uranium in the Environment: Behavior and Toxicity. Figure 3

Uranium uptake after 30 min by the unicellular green alga *Chlamydomonas reinhardtii* (Modified from Fortin et al. [32])

Uranium in the Environment: Behavior and Toxicity.

Table 6 Optimum pH for uptake of uranium by several algal species (Kalin et al. [7])

Taxa	Optimum pH
<i>Anacysti</i> (Cyanobacterium)	3.0 and 5.0
<i>Sargassum</i> (Phaeophyta)	2.5 and 4.0
<i>Scenedesmus</i> (Chlorophyta)	5.0 and 8.5
<i>Scenedesmus</i> (Chlorophyta)	5.9 and 6.8
<i>Chlorella regularis</i> (Chlorophyta)	6.0 and 7.0

eightfold increase in both intracellular and extracellular U at both pH 5.7 and 6.5. Intracellular U concentrations were greater at pH 6.5 than at pH 5.7 with about two times the U located in the cell. Greater toxicity at pH 6.5 corresponded with greater uptake of U to the cell surface and uptake of U into the cell. Speciation modeling indicates that the decreasing toxicity of U to *Chlorella* sp. with decreasing pH may have been the result of H^+ inhibiting intracellular U uptake at the lower pH. At the pH range 5.7–6.5, $\text{UO}_2(\text{OH})_3\text{CO}_3^-$ is expected to dominate (70–93%) and <4% would occur as UO_2^{2+} [5].

Charles et al. [163] assessed the effect of hardness at 8, 40, 100, and 400 $\text{mg}\cdot\text{L}^{-1}$ as CaCO_3 on the 72-h growth inhibition of *Chlorella* sp. A 50-fold increase in hardness resulted in a fivefold decrease in U toxicity, from 56 to 270 $\mu\text{g}\cdot\text{L}^{-1}$. Reduction in toxicity was attributed to competition with calcium and magnesium for binding sites. The effect of hardness on extracellular and intracellular U concentrations, and U toxicity is summarized in Table 5. Intracellular U concentrations decreased with increasing hardness, whereas the ratio of extracellular to intracellular U increased with increasing U concentration. Hardness also reduced the toxicity of U to *Chlorella* sp (Table 7). The toxicity of U to *Pseudokirchneriella subcapitata* (previously *Selenastrum capricornutum*) was in general agreement with the findings for *Chlorella*. The alga *P. subcapitata* had $\text{IC}_{25\text{S}}$ for inhibition of growth of 27–150 $\mu\text{g}\cdot\text{L}^{-1}$ depending on hardness (Table 5) [147].

Macrophytes

Van Dam et al. [166] investigated the toxicity of U to the tropical duckweed *Lemna aequinoctialis*. An IC_{50} of > 2,850 $\mu\text{g}\cdot\text{L}^{-1}$ U for growth inhibition and an NOEC of 250 $\mu\text{g}\cdot\text{L}^{-1}$ were reported for *L. aequinoctialis*. These values indicate that duckweed is not very sensitive to U.

Uranium in the Environment: Behavior and Toxicity.**Table 7** Effect of hardness ($\text{mg}\cdot\text{L}^{-1}$ CaCO_3) on uranium uptake by *Chlorella* sp. (From Charles et al. [163])

Hardness	$\mu\text{g U}\cdot\text{L}^{-1}$	Intracellular $(\times 10^{-6} \text{ ng } \mu\text{m}^{-3})$	Extracellular $\text{U } (\times 10^{-6} \text{ ng } \mu\text{m}^{-2})$
8	12	0.60	0.41
	53	0.82	0.80
	270	4.6	15
400	17	0.07	0.03
	62	0.10	0.22
	310	0.15	0.36
	520	1.1	5.0

This is in agreement with the findings for *Lemna minor*, which had a 7-day IC_{50} of $7,400 \mu\text{g}\cdot\text{L}^{-1}$ and 7-day IC_{25} of $4,700 \mu\text{g}\cdot\text{L}^{-1}$ for number of fronds and IC_{50} of $13,100 \mu\text{g}\cdot\text{L}^{-1}$ and IC_{25} of $6,400 \mu\text{g}\cdot\text{L}^{-1}$ for growth at low hardness [147].

Benthic Invertebrates

Sediments are a sink for particle-reactive contaminants released into aquatic ecosystems. Sediment-associated contaminants may have adverse effects on benthic macroinvertebrates and fish that feed on contaminated invertebrates. Borgmann et al. [167] reported a 7-day LC_{50} of $21 \mu\text{g}\cdot\text{L}^{-1}$ (measured) for *H. azteca* in soft water (hardness $18 \text{ mg}\cdot\text{L}^{-1}$) and a LC_{50} of $1,651 \mu\text{g}\cdot\text{L}^{-1}$ (nominal) in hard water (hardness $124 \text{ mg}\cdot\text{L}^{-1}$). This is in agreement with the 14-day LC_{50} of $17 \mu\text{g}\cdot\text{L}^{-1}$ reported for *H. azteca* in soft water by Vizon SciTec [147] and the 14-day LC_{50} reported for *H. azteca* of $1.52 \text{ mg}\cdot\text{L}^{-1}$ in hard water by Kuhne et al. [159]. A pronounced hardness effect was seen in the Vizon SciTec [147] study.

When *H. azteca* was exposed to U spiked sediment, it was concluded that toxicity was mainly from the water rather than the sediment [168]. The pH of the water (pH 6.9–8.0) had a marked effect on the uptake of U by *H. azteca*, whereas calcium concentration (4, 30 or 40 mg) had little effect on U uptake but did influence U speciation in aqueous solution. The 28-day LC_{50} s for juvenile *H. azteca* ranged from 14.5 to $> 714 \mu\text{g}\cdot\text{L}^{-1}$ (4.6 – $14 \text{ mg}\cdot\text{g}^{-1}$ dry sediment) depending on pH and

calcium concentration (Table 5), whereas 28-day LC_{50} s ranged from 15 to $2,060 \mu\text{g}\cdot\text{L}^{-1}$ (5.5 – $13.1 \text{ mg U}\cdot\text{g}^{-1}$ dry sediment) for adult specimens (Table 5). BEAK International Inc. [177] reported juvenile *H. azteca* were much more sensitive to U than the adults with a LC_{20} of $15 \text{ mg U}\cdot\text{kg}^{-1}$ dw for juveniles compared to a LC_{20} of $116 \text{ mg U}\cdot\text{kg}^{-1}$ dw for adults in hard water (105 – $308 \text{ mg}\cdot\text{L}^{-1}$ as CaCO_3).

The midge *Chironomus tentans* was found to be quite tolerant to U with a 10-day LC_{50} of $6,400 \mu\text{g}\cdot\text{L}^{-1}$ and a LOEC of $1,519 \mu\text{g}\cdot\text{L}^{-1}$ [169]. Survival was a more sensitive endpoint than growth. In contrast, *Chironomus riparius* appears to be quite sensitive to U. First-instar larvae of *C. riparius* were exposed to sediment U concentrations up to $23.84 \text{ mg U}\cdot\text{kg}^{-1}$ dw in 10-day bioassays, and associated aqueous U concentrations up to $85.17 \mu\text{g}\cdot\text{L}^{-1}$ as a result of desorption of U from the sediment [170]. Mortality of larvae was higher at sediment U concentrations of $6.07 \text{ mg U}\cdot\text{kg}^{-1}$ dw, than $2.97 \text{ mg U}\cdot\text{kg}^{-1}$ dw sediments and the control. Growth of fourth instar larvae was significantly lower at $2.97 \text{ mg U}\cdot\text{kg}^{-1}$ dw sediment than in control sediment. No larvae matured to the fourth instar stage over the 10-day study in the $23.84 \mu\text{g U}\cdot\text{kg}^{-1}$ dw sediment. However, desorption of U from sediments resulted in aqueous U concentrations of 10 , 28 , 51 , and $85 \mu\text{g}\cdot\text{L}^{-1}$, respectively (Table 5), and it can be speculated that the toxicity was mainly from the water pathway as reported by Alves et al. [168].

The freshwater snail *Amerianna cumingi* was found to be sensitive to U with an IC_{50} of $250 \mu\text{g}\cdot\text{L}^{-1}$ and an IC_{10} of $22 \mu\text{g}\cdot\text{L}^{-1}$ for inhibition of egg production [165]. The Asiatic clam (*Corbicula fluminea*) was more sensitive to U at pH 5.5 than at pH 6.5. The minimal sensitivity threshold that caused closure of the valve by 50% of the bivalves was $12 \mu\text{g}\cdot\text{L}^{-1}$ at pH 5.5 after 5 h of exposure. Higher concentrations of UO_2^{2+} were required at pH 6.5 to illicit the same response. In both marine and freshwater mollusks and crustaceans, U accumulates in lysosomes [179].

In another investigation using the mayfly *Hexagenia limbata* nymphs exposed to contaminated sediment from historic uranium mining operations, Jaagumagi and Bedard [180] reported a 63% mortality rate in nymphs exposed to $491 \text{ mg U}\cdot\text{kg}^{-1}$ dw sediment

in 22-day sediment bioassays. No effect on mortality or growth was observed for mayfly nymphs at U sediment concentrations of 78, 83, and 100 mg·kg⁻¹ dw.

From the above it can be concluded that U can be toxic at low concentrations to certain aquatic organisms under specific conditions such as low alkalinity.

Overview of Parameters Affecting U Toxicity

Hardness Hardness is reported to reduce the toxicity of uranium to freshwater organisms [148, 149, 157]. However, a quantitative relationship describing U toxicity as a function of hardness is lacking. This is because many experiments designed to investigate the effects of hardness on U toxicity have confounded the effects of true water hardness (calcium or magnesium concentrations) with alkalinity (carbonate concentrations) and pH (H⁺ concentration) [35].

Hardness (calcium or magnesium concentrations) and alkalinity (carbonate concentration) influence toxicity through different mechanisms. In the case of hardness, calcium and/or magnesium competitively inhibit the uptake at the cell surface of trace metals thereby reducing their toxicity. About sevenfold more Mg is needed to exhibit the same ameliorative effect as Ca on toxicity [181, 182]. In the case of alkalinity, complexation of trace metals with carbonate modifies metal speciation and reduces the concentration of the toxic metal species, UO₂ and UO₂OH⁺ [34].

In studies on the effect of hardness on U toxicity in which CaSO₄ was added to prevent alkalinity from covarying with hardness, hardness was found to decrease the toxicity of U to the alga *Chlorella* sp. [163] and the macrophyte *L. minor* [147]. A 50-fold (mg·L⁻¹) increase in hardness resulted in a 4.8-fold (mg·L⁻¹) decrease in U toxicity [163]. In contrast, U toxicity to the alga *P. subcapitata* [147] was not influenced by hardness. The invertebrate *H. azteca* responded markedly to hardness; a 16-fold increase in hardness decreased U toxicity by 12-fold in 14-day toxicity tests [147] (Table 5). Other studies that increased hardness with alkalinity also reported reduced toxicity, for example, to *D. magna* [36, 157] and the green hydra *Hydra viridissima* [35].

Equations have been developed to predict the effect of hardness on U toxicity to aquatic organisms [100, 157]. However, when EC [25] plotted hardness versus

LC₅₀ data similar to that described by Sheppard et al. [100], but included more recent data that did not covary with alkalinity [147], they found no significant relationship with which to predict U toxicity with hardness.

Not all organisms show a hardness effect in their response to U toxicity. For example, for rainbow trout and fathead minnow, U toxicity did not vary in short-term tests where alkalinity was held constant as hardness increased [147]. Similar results were found in long-term tests on *C. dubia* [147]. Therefore, in setting benchmarks to protect aquatic biota the hardness should not be a consideration if the goal is to protect most species.

Alkalinity Only two toxicity studies were found that altered alkalinity at constant hardness. Riethmuller et al. [35] varied alkalinity from 4.0 to 102 mg CaCO₃·L⁻¹ to assess U toxicity to the freshwater hydrazoan *Hydra viridissima*. Increased alkalinity changed the speciation of U (percent UO₂(CO₃)₂²⁻ from <1% to 20% and UO₂²⁺ from 6% to 1%) but had no effect on U toxicity. Markich et al. [34] reported toxicity was inversely proportional to alkalinity when pH and hardness were kept constant. U toxicity decreased 20% with a fivefold increase in alkalinity at pH 5 and hardness of 3.5 mg·L⁻¹ as CaCO₃.

There is not enough information on the effect of alkalinity on U toxicity, and further studies are required with a variety of organisms to establish a relationship.

Natural Organic Matter Natural organic matter has the potential to bind the toxic free ion forms of a metal and hence reduce toxicity similar to alkalinity. Markich et al. [29] found that an increase in fulvic acid from 0 to 7.91 mg·L⁻¹ decreased U toxicity to a freshwater bivalve by a factor of 2–3. In another study, Hogan et al. [164] found increasing dissolved organic matter decreased U toxicity in *Chlorella* sp. More work is required to assess the effect of organic matter on U toxicity.

Temperature Water temperature may affect U toxicity by changing U solubility, speciation or kinetics, and the metabolic rate of organisms, for example, uptake rate, depuration rate, and survival. Information could not be found on the effect of temperature on U speciation or

toxicity. However, temperature affects the metabolism of poikilotherms, which will affect toxicity, for example, a 10°C increase in temperature may increase metabolism by factor of 2 and toxicity by a factor of 2 or 3 [183]. Transfer of U to organic sediments is also more rapid at higher temperature [69].

Metal Mixtures Only one study examined the toxicity of mixtures of U and other metals. Hamilton and Buhl [169] exposed the flannelmouth sucker, *C. latipinnis*, to metal mixtures consisting of copper, zinc, selenium, boron, vanadium, uranium, and arsenic, based on concentrations found at their study site, the San Jan River, USA. Toxicity of metals was found to be additive. Further work needs to be done in this area, especially since mine and mill effluents contain elevated levels of various metals, including U.

Regulatory Guidelines Regulatory guidelines are available from a number of jurisdictions (Table 8). Until recently, the Canadian Water Quality Guidelines for Protection of Aquatic Life and Wildlife was 300 $\mu\text{g}\cdot\text{L}^{-1}$ [8]. Recently Environment Canada prepared a draft guideline which is undergoing review. The proposed guideline is considerably lower than their former guideline and is more in line with those derived independently by several Canadian provinces. However, it is considerably higher than the guideline of 0.5 $\mu\text{g}\cdot\text{L}^{-1}$ for Australia and New Zealand [184], which includes an

application factor of 20 because of the limited amount of long-term data available for tropical species.

Future Directions

Studies demonstrate that background concentrations of U found in the environment are unlikely to be toxic to biota, and if toxicity occurs, it is more likely to be chemotoxic than radiotoxic because of the low specific activity of ^{238}U . Uranium usually has low toxicity primarily because oral uptake rates are low and concentrations decrease up the food chain. However, U may be toxic to organisms in areas with elevated U concentrations above background levels.

The toxicity of U to human and terrestrial wildlife has been well documented based on human epidemiological studies and laboratory studies on mammals. Further work is required on the toxicity of U to terrestrial plants and soil invertebrates, as few studies have focused on these organisms.

A substantial amount of work has been undertaken on the effects of U on the aquatic ecosystem. Nevertheless, there remain many unanswered questions because of the complex chemistry of U and the interactions between U and environmental parameters. For aquatic organisms, U toxicity for several species decreases with an increase in hardness and alkalinity. However, not all organisms show a hardness effect on toxicity, and U may be toxic at low concentrations as observed for fish (*M. mogurnda* and early life stages of rainbow trout), zooplankton (*C. dubia*, *D. pulex*, and *D. magna*), algal spp. (*Chlorella* sp. and *P. subcapitata*), and benthic invertebrates (*H. azteca*, *C. riparius*, *C. fluminea*, and *H. viridissima*) under the conditions given in Table 5. Other parameters, such as pH, alkalinity, and ligands affect U speciation and their effect on toxicity needs assessment at environmentally relevant concentrations and conditions. In the case of aqueous releases from mining/milling operations, the interactions of the ameliorating effects of elevated calcium and magnesium concentrations from effluent treatment, elevated U concentrations with potentially increased abundance of toxic U species (UO_2^{2+} and UO_2OH^+) along with increased concentrations of sulfate and other metals must be considered. Site-specific testing of the receiving surface waters is required to assess the potential toxicity of complex effluents from mining/milling U operations.

Uranium in the Environment: Behavior and Toxicity.

Table 8 Water Quality Guidelines for Protection of Aquatic life from U

Jurisdiction	U concentration ($\mu\text{g}\cdot\text{L}^{-1}$)	Hardness ($\text{mg}\cdot\text{L}^{-1} \text{CaCO}_3$)
Quebec	14	20–100
	100	100–210
Saskatchewan	15	
Ontario	5	
China	50	
Australia	0.5	
New Zealand	0.5	
Canada	300 (being revised)	

Bibliography

Primary Literature

- Bleise A, Danesi PR, Burkart W (2003) Properties, use and health effects of depleted uranium (DU): a general overview. *J Environ Radioact* 64:93–112
- Ribera D, Labrot F, Tisnerat G, Narbonne JF (1996) Uranium in the environment: occurrence, transfer, and biological effects. *Rev Environ Contam Toxicol* 146:53–89
- Meinrath A, Schneider P, Meinrath G (2003) Uranium ores and depleted uranium in the environment, with a reference to uranium in the biosphere from the Erzgebirge/Sachsen, Germany. *J Environ Radioact* 64:175–193
- Harley NH (1996) Toxic effects of radiation and radioactive materials. In: Klaassen CD (ed) Casarett and Doull's toxicology: the basic science of poisons. McGraw-Hill, New York, pp 773–800
- Markich S (2002) Uranium speciation and bioavailability in aquatic systems: an overview. *Sci World J* 2:707–729
- Environmental Canada, Health Canada (2003) Canadian Environmental Protection Act 1999. Priority substances list assessment report releases of radionuclides from nuclear facilities (Impact on non-human biota). Final Report May 2003
- Kalin M, Wheeler WN, Meinrath G (2005) The removal of uranium from mining waste water using algal/microbial biomass. *J Radiochem Radioact* 78:151–177
- Environment Canada (1983) Guideline for surface water quality volume 1: inorganic chemical substances URANIUM. Inland Waters Directorate, Water Quality Branch, Ottawa
- Painter S, Cameron EM, Allan R, Rouse J (1994) Reconnaissance geochemistry and its environmental relevance. *J Geochem Explor* 51:213–246
- CCME (Canadian Council of Ministers of the Environment) (2007) Canadian soil quality guidelines for uranium: environmental and human health scientific supporting document PN 1371 ISBN 978-1-896997-64-3 PDF
- Gascoyne M (1992) Geochemistry of the actinides and their daughters. In: Ivanovich M, Harmon RS (eds) Uranium series disequilibrium: applications to earth, marine and environmental sciences, 2nd edn. Clarendon, Oxford, pp 34–61
- Macdonald CR (1998) The chemical toxicity of uranium to wildlife. Report to the Commercial Chemical Division, Environment Canada. Contract Number K2221-7-0076. Pinawa, MB. 32 pp
- Ragnarsdottir KV, Charlet L (2000) Uranium behavior in natural environments. In: Cotter-Howells JD, Campbell LS, Valsami-Jones E, Batchelder M (eds) Environmental mineralogy: microbial interactions, anthropogenic influences, contaminated land and waste management, vol 9, Mineralogical Society Series. Mineralogical Society, London, pp 245–289
- Sheppard MI (1980) The environmental behaviour of uranium and thorium. Whiteshell Nuclear Research Establishment, Atomic Energy of Canada Limited, Pinawa, Manitoba, 44 pp. (Technical Report AECL-6795)
- Fernandes HM, Filho FFLS, Perez V, Franklin MR, Gomiero LA (2006) Radioecological characterization of a uranium mining site located in a semi-arid region of Brazil. *J Environ Radioact* 88:140–157
- Landa ER, Gray JR (1995) US Geological Survey research on the environmental fate of uranium mining and milling wastes. *Environ Geol* 26:19–31
- Joshi SR, Waite DT, Platford RF (1989) Vertical distribution of uranium mill tailings contaminants in Langley Bay, Lake Athabasca sediments. *Sci Tot Environ* 87(88):85–104
- Thomas PA (1997) The ecological distribution and bioavailability of uranium series radionuclides in terrestrial food chains: key Lake uranium operations, northern Saskatchewan. Prepared for Environmental Protection, Prairie and Northern Region, Environment Canada, Regina, Saskatchewan, by P.A. Thomas, University of Saskatchewan, Saskatoon, Saskatchewan
- Robertson EL, Liber K (2007) Bioassays with caged *Hyalella azteca* to determine in situ toxicity downstream of two Saskatchewan, Canada, uranium operations. *Environ Toxicol Chem* 26:2345–2355
- Shropshire DE, Williams KA, Boore WB, Smith JD, Dixon BW, Dunzik-Gougar M, Adams RD, Combert D, Schneider E (2008) Advanced fuel cycle cost basis. Report prepared for Idaho National laboratory, INL/EXT-07-12107 Rev. 1
- Diehl P (1999) Depleted uranium: a by-product of the nuclear chain. <http://www.antenna.nl/wise-database/uranium/dhap99f.html>. Accessed 22 Nov 2009
- Tracy BL, Prantl FA (1985) Radiological impact of coal-fired power generation. *J Environ Radioact* 2:145–160
- Chen JH, Edwards RL, Wasserburg GJ (1986) ^{238}U , ^{234}U & ^{232}Th in seawater. *Earth Planet Sci Lett* 80:241–251
- Choppin GR, Stout BE (1989) Actinide behaviour in natural waters. *Sci Total Environ* 83:203–216
- Environment Canada (2008) Canadian water quality guidelines for uranium: Scientific supporting document (Draft) 10th March, 2004 original draft updated 5 Feb 2008
- Bosshard E, Zimmerli B, Schlatter Ch (1992) Uranium in the diet: risk assessment of its nephro- and radiotoxicity. *Chemosphere* 24:309–321
- Langmuir D (1978) Uranium solution-mineral equilibria at low temperatures with application to sedimentary ore deposits. *Geochim Cosmochim Acta* 42:547–569
- Hamelink JL, Landrum PF, Bergman HL, Benson WH (1994) Bioavailability: physical, chemical and biological interactions. Lewis, Boca Raton
- Markich SJ, Brown PL, Jeffree RA, Lim RP (2000) Valve movement responses of *Velesunio angasi* (Bivalvia: Hyriidae) to manganese and uranium: an exception to the free ion activity model. *Aquat Toxicol* 1:155–175
- Denison FH, Garnier-Laplace J (2005) The effects of database parameter uncertainty on uranium (VI) equilibrium calculations. *Geochim Cosmochim Acta* 69:2183–2191
- Unsworth ER, Jones P (2002) The effect of thermodynamic data on computer model predictions of uranium speciation in natural water systems. *J Environ Monit* 4:528–532
- Fortin C, Dutel L, Garnier-Laplace J (2004) Uranium complexation and uptake by green alga in relation to chemical

- speciation: the importance of the free uranyl ion. *Environ Toxicol Chem* 23:974–981
33. Fournier ED, Tran F, Denison J, Massabuau C, Garnier-Laplace J (2004) Valve closure response to uranium exposure for a freshwater bivalve (*Corbicula fluminea*): qualification of the influence of pH. *Environ Toxicol Chem* 23:1108–1114
 34. Markich SJ, Brown PL, Jeffree RA (1996) The use of geochemical speciation modelling to predict the impact of uranium to freshwater biota. *Radiochemica Acta* 74:321–326
 35. Reithmuller N, Markich SJ, Van Dam RA, Parry D (2001) Effects of water hardness and alkalinity on the toxicity of uranium to a tropical freshwater hydra (*Hydra viridissima*). *Biomarkers* 6:45–51
 36. Barata C, Baird DJ, Markich SJ (1998) Influence of genetic and environmental factors on the tolerance of *Daphnia magna* straus to essential and non-essential metals. *Aquat Toxicol* 42:115–137
 37. Higgo JJW (1987) Clay as a barrier to radionuclide migration. *Progr Nucl Energy* 19:173–207
 38. Bernhard G, Geipel G, Brendler V, Nitsche H (1998) Uranium speciation in waters of different uranium mining areas. *J Alloy Comp* 271–273:201–205
 39. Melville GE (1995) Changes in the pelagic crustacean zooplankton of high-boreal Island Lake, Saskatchewan, associated with uranium mining. *Environ Monit Assess* 34:45–58
 40. Sheppard SC (1995) Parameter values to model the soil ingestion pathway. *Environ Monit Assess* 34:27–44
 41. Fleming S, Marsh M, Wagenaar A, McLaughlin D, Bloxam R (2000) Proposed ambient air quality standards for uranium. Rational Document: for discussion with Port Hope stakeholders. Ontario Ministry of the Environment
 42. Sheppard SC, Evenden WG (1992) Bioavailability indices for uranium: effect of concentration in eleven soils. *Arch Environ Contam Toxicol* 23:117–124
 43. Batson VL, Bertsch PM, Herbert BE (1996) Transport of anthropogenic uranium from sediments to surface waters during episodic storm events. *J Environ Qual* 25:1129–1137
 44. Babich H, Stotzky G (1985) Heavy metal toxicity to microbe-mediated ecological processes: a review and potential application to regulatory policies. *Environ Res* 36:111–137
 45. Coughtrey PJ, Jones CH, Martin MH, Shales SW (1979) Litter accumulation in woodlands contaminated by Pb, Zn, Cd and Cu. *Oecologia* 39:51–60
 46. Bird GA, Stephenson M, Cornett RJ (1993) The surface water sub model for the assessment of Canada's Nuclear fuel waste management concept. AECL research report AECL-10290, COG-91-193. Pinawa, Manitoba
 47. Bird GA, Schwartz W (1996) Nuclide concentration factors for freshwater biota. Atomic Energy of Canada limited, Pinawa, Manitoba (Technical Record TR-703, COG-95-397)
 48. FASSET Frederica radiation effects database federica-online.org <http://www.frederica-online.org/panel.asp>. Accessed 22 Nov 2009
 49. International Atomic Energy Agency (1985) Radionuclides in the marine environment. Technical Reports Series No. 247
 50. International Atomic Energy Agency (1994) Handbook of parameter values for the prediction of radionuclide transfer in temperate environments. Tech Report Series No 364
 51. International Atomic Energy Agency (2010) Modeling the environmental transfer of tritium and carbon-14 to biota and man. Final report, Tritium and Carbon-14 Working Group, IAEA EMRAS Program, International Atomic Energy Agency, Vienna Australia
 52. Sheppard MI, Thibault DH (1990) Default soil solid/liquid partition coefficients, K_d s, for four major soil types: a compendium. *Health Phys* 59:471–482
 53. Kumblad L, Kautsky U, Naeslund B (2006) Transport and fate of radionuclides in aquatic environments – the use of ecosystem modelling for exposure assessments of nuclear facilities. *J Environ Radioact* 87:107–129
 54. Sheppard MI, Thibault DH (1984) Natural uranium concentrations of native plants over a low-grade ore body. *Can J Bot* 62:1069–1075
 55. Mahon DC (1982) Uptake and translocation of naturally-occurring radionuclides of the uranium series. *Bull Environ Contam Toxicol* 29:697–703
 56. Hanson WC, Miera FR Jr, (1976) Long-term ecological effects of exposure to uranium. Technical Report LA-6269 Los Alamos Scientific Lab, New Mexico, USA
 57. Beyer WN, Connor EE, Gerould S (1994) Estimates of soil ingestion by wildlife. *J Wildl Manage* 58:375–382
 58. Linsalata P, Morse R, Ford H, Eisenbud M, Franca EP, de Castro MB, Lobao N, Sachett I, Carlos M (1989) Transport pathways of Th, U, Ra and La from soil to cattle tissues. *J Environ Radioact* 10:115–140
 59. Weeks HP, Kirkpatrick CM (1976) Adaptations of white-tailed deer to naturally occurring sodium deficiencies. *J Wildl Manage* 40:610–625
 60. Sullivan MF (1980) Absorption of actinide elements from the gastrointestinal tract of neonatal animals. *Health Phys* 38:173–185
 61. Hesslein RH (1987) Whole-lake metal radiotracer movement in fertilized lake basins. *Can J Fish Aquat Sci* 44(suppl 1): 74–82
 62. Fisher NS, Teyssie JL, Krishnaswami S, Baskaran M (1987) Accumulation of Th, Pb, U, and Ra in marine phytoplankton and its geochemical significance. *Limnol Oceanogr* 32:131–142
 63. Degens ET, von Bronsart G, Ho Kim W, Khoo F, Dickman MD (1979) Environmental parameters responsible for the fixation of uranium in recent sediments; test area Bow lake, Ontario, Canada, Part 1. *Mitteilungen aus dem Geologisch-Palaeontologischen Institut der Universitaet Hamburg*, no. 49, pp 27–60
 64. Nyffeler UP, Santschi PH, Li YH (1986) The relevance of scavenging kinetics to modelling of sediment-water interactions in natural waters. *Limnol Oceanogr* 31:277–292
 65. Bird GA, Stephenson M, Cornett RJ (1993) The surface water model for assessing Canada's nuclear fuel waste disposal concept. *Waste Manag* 13:153–170

66. Santschi PH, Nyffeler UP, Anderson RF, Schiff SL, O'Hara P, Hesslein RH (1986) Response of radioactive trace metals to acid-base titrations in controlled experimental ecosystems: evaluation of transport parameters for application to whole-lake radiotracer experiments. *Can J Fish Aquat Sci* 43:60–77
67. Carignan R, Nriagu JO (1985) Trace metal deposition and mobility in the sediments of two lakes near Sudbury, Ontario. *Geochim Cosmochim Acta* 49:1753–1764
68. Swanson SM (1985) Food-chain transfer of U-series radionuclides in a northern Saskatchewan aquatic system. *Health Phys* 49:747–770
69. Bird GA, Evenden WG (1994) Effect of sediment type, temperature and colloids on the transfer of radionuclides from water to sediment. *J Environ Radioact* 22:219–242
70. Ames LL, McGarrah JE, Walker BA (1983) Sorption of uranium and radium by biotite, muscovite, and phlogopite. *Clays Clay Mater* 31:343–351
71. Santschi PH, Honeyman BD (1989) Radionuclides in aquatic environments. *Radiat Phys Chem* 34:213–40
72. Livens FR (1991) Chemical reactions of metals with humic material. *Environ Pollut* 70:183–208
73. Bird GA, Schwartz WJ, Motycka M (1998) Fate of ^{60}Co and ^{134}Cs added to the hypolimnion of a Canadian Shield lake: accumulation in biota. *Can J Fish Aquat Sci* 55:987–998
74. Likens GE, Davis MB (1975) Post-glacial history of Mirror Lake and its watershed in New Hampshire, U.S.A. An initial report. *Int Ver Theor Angew Limnol* 19:982–993
75. Evans RD, Rigler FH (1983) A test of the lead-210 dating for the measurement of whole lake soft sediment accumulation. *Can J Fish Aquat Sci* 40:506–515
76. Shaw TJ, Sholkovitz ER, Kinkhamer G (1994) Redox dynamics in the Chesapeake Bay; The effect on sediment/water uranium exchange. *Geochim Cosmochim Acta* 58:2985–2995
77. Alberic P, Viollier EP, Jezequel D, Grosbois C, Michard G (2000) Interactions between trace elements and dissolved organic matter in the stagnant anoxic deep layer of a meromictic lake. *Limnol Oceanogr* 45:1088–1096
78. Edgington DN, Robbins JA, Colman SM, Orlandini KA, Gustin MP (1996) Uranium-series disequilibrium, sedimentation, diatom fustules, and paleoclimate change in Lake Baikal. *Earth Plant Sci Lett* 142:29–42
79. Hart BT (1982) Uptake of trace metals by sediments and suspended particulate: a review. *Hydrobiologia* 91:299–313
80. Alberts JJ, Wahlgren MA, Orlandini KA, Durbahn CA (1989) The distributions of $^{239,240}\text{Pu}$, ^{238}Pu , ^{241}Am and ^{137}Cs among chemically-defined components of sediments, settling particulates and net plankton of lake Michigan. *J Environ Radioact* 9:89–103
81. Jenne EA (1968) Controls on Mn, Fe, Co, Ni, Cu, and Zn concentrations in soils and water: the significant role of hydrous Mn and Fe oxides. In: Gould RF (ed) *Trace inorganics in water*, vol 73. American Chemical Society, Washington DC, pp 337–387
82. Wetzel RG (1975) *Limnology*. WB Saunders, Philadelphia
83. Kovalsky VV, Vorotnitskaya IE (1965) Biogenic migration of uranium in Lake Issyk-Kul'. *Geokhimiya* 6:724–732
84. Otton JK, Zielinski RA, Been JM (1989) Uranium in Holocene valley-fill sediments, and uranium, radon and helium in waters, Lake Tahoe-Carson Range Area Nevada and California. *U.S.A. Environ Geol Water Sci* 13:15–28
85. Landa ER (1982) Leaching of radionuclides from uranium ore and mill tailings. *Uranium* 1:53–64
86. Bird GA, Schwartz WJ, Motycka M, Rosentreter J (1998) Behaviour of ^{60}Co and ^{134}Cs in a Canadian shield lake over five years. *Sci Tot Environ* 212:115–135
87. Simpson SL, Apte SC, Batley GE (1998) Effect of short-term resuspension events on trace metal speciation in polluted anoxic sediment. *Environ Sci Technol* 32:620–625
88. Walgren MA, Alberts JJ, Nelson DM, Orlandini KA (1976) Study of the behaviour of transuranics and possible chemical analogues in Lake Michigan water and biota. In: IAEA, *Transuranium nuclides in the environment*. Symposium Proceedings, San Francisco, pp 9–24. STI/PUB/410, International Atomic Energy Agency, Vienna
89. Swanson SM (1983) Levels of ^{226}Ra , ^{210}Pb , and ^{238}U in fish near a Saskatchewan uranium mine and mill. *Health Phys* 45:67–80
90. Ahsanullah M, Williams AR (1989) Kinetics of uranium uptake by the crab *Pachygrapsus laevis* and the zebra winkle *Austrocochlea constricta*. *Mar Biol* 101:323–327
91. Mirka MA, Clulow FV, Dave NK, Lim TP (1986) Radium-226 in cattails, *Typha latifolia* and bone of muskrat, *Ondatra zibethica* (L.) from a watershed with uranium tailings near the city of Elliot Lake, Canada. *Environ Pollut* 91:41–51
92. Waite DT, Joshi SR, Sommerstad H (1988) The effect of uranium mine tailings on radionuclide concentrations in Langley Bay, Saskatchewan, Canada. *Arch Environ Contam Toxicol* 17:373–380
93. Golder Associates Ltd. (2002) Final report on current period environmental monitoring program for the Beaverlodge Mine site in northern Saskatchewan. Submitted to Cameco Corporation, Saskatoon, Saskatchewan, by Golder Associates Ltd., Saskatoon, Saskatchewan
94. Aery NC, Jain GS (1998) Influence of uranium on the growth of wheat. *J Environ Biol* 19:5–23
95. Gulati KL, Oswal MC, Nagpaul KK (1980) Assimilation of uranium by wheat and tomato plants. *Plant Soil* 55:55–59
96. Meyer MC, Paschke MW, McLendon T, Price D (1998) Decreases in soil microbial function and functional diversity in response to depleted uranium. *J Environ Qual* 27:1306–1311
97. Sheppard SC (1989) Toxicity levels of arsenic and uranium in soils. Atomic Energy of Canada Limited, Pinawa, Manitoba (AECL Technical Report TR-480)
98. Sheppard MI, Thibault DH, Sheppard SC (1985) Concentrations and concentration ratios of U, As and Co in Scots pine grown in waste-site soil and an experimentally contaminated soil. *Water Air Soil Pollut* 26:85–94
99. Sheppard SC, Evenden WG, Anderson AJ (1992) Multiple assays of uranium toxicity in soil. *Environ Toxicol Wat Qual* 7:275–294

100. Sheppard SC, Sheppard MI, Gallerand M-O, Sanipelli B (2005) Derivation of ecotoxicity threshold for uranium. *J Environ Radioact* 79:55–83
101. Weinberger P, Murthy TCSS (1985) Uranium uptake and its effects on the germination and early growth of some crop species. Report for Atomic Energy of Canada Limited, Whiteshell Nuclear Research Establishment, Pinawa, Manitoba Technical Record TR-329
102. Kabata-Pendias A, Pendias H (1992) Trace elements in soils and plants, 2nd edn. CRC Press, Boca Raton, 365 pp
103. Durbin PW, Wrenn ME (1975) Metabolism and effects of uranium in animals. Conference on occupational health experience with uranium. Arlington, Virginia, April 28–30, 1975. U.S. Energy Research & Development Administration; U.S. Government Printing Office, Washington, DC, pp 68–129 (ERDA 93)
104. Leach LJ, Yuile CL, Hodge HC, Sylvester GE, Wilson HB (1973) A five-year inhalation study with natural uranium dioxide (UO_2) dust –II. Postexposure retention and biological effects in the monkey, dog and rat. *Health Phys* 25:239–258
105. Leach LJ, Maynard EA, Hodge HC, Scott JK, Yuile CL, Sylvester GE, Wilson HB (1970) A five year inhalation study with natural uranium dioxide (UO_2) dust –I. Retention and biological effect in the monkey, dog and rat. *Health Phys* 18:599–612
106. Kocher DC (1989) Relationship between kidney burden and radiation dose from chronic ingestion of U: implications for radiation standards for the public. *Health Phys* 57:9–15
107. Yuile CL (1973) Animal experiments. In: Hodge HC, Stannard JN, Hursh JB (eds) Uranium, plutonium, transplutonic elements. Springer, New York, pp 165–196
108. Llobet JM, Sirvent JJ, Ortega A, Domingo JL (1991) Influence of chronic exposure to uranium on male reproduction in mice. *Fund Appl Toxicol* 16:821–829
109. Harrison JD (1991) The gastrointestinal absorption of the actinide elements. *Sci Tot Environ* 100:43–60
110. Wrenn ME, Durbin PW, Howard B, Lipsztein J, Rundo J, Still ET, Willis DL (1985) Metabolism of ingested U and Ra. *Health Phys* 48:601–633
111. Wrenn ME, Bertelli L, Durbin PW, Eckerman KF, Lipsztein JL, Singh NP (1995) A biokinetic and dosimetric model for the metabolism of uranium. Atomic Energy Control Board, Ottawa, Ontario 79 pp (INFO-0500)
112. International Commission on Radiological Protection (1979) Limits on intake of radionuclides by workers. ICRP Publication 30, Part 1. Pergamon, New York
113. Leggett RW, Harrison JD (1995) Fractional absorption of ingested uranium in humans. *Health Phys* 68:484–498
114. Moss MM (1989) Study of the effects of uranium on kidney function. Atomic Energy Control Board INFO-0306
115. Sullivan MF, Gorham LS (1982) Further studies on the absorption of actinide elements from the gastrointestinal tract of neonatal animals. *Health Phys* 43:509–519
116. LaTouche YD, Willis DL, Dawydiak OI (1987) Absorption and biokinetics of U in rats following oral administration of uranyl nitrate solution. *Health Phys* 53:147–162
117. Tracy BL, Quinn JM, Lahey J, Gilman AP, Mancuso K, Yagminas AP, Villeneuve DC (1992) Absorption and retention of uranium from drinking water by rats and rabbits. *Health Phys* 62:65–73
118. Morris SC, Meinhold AF (1995) Probabilistic risk assessment of nephrotoxic effect of uranium in drinking water. *Health Phys* 69:897–908
119. Hursh JB, Neuman WF, Toribara T, Wilson H, Waterhouse C (1969) Oral ingestion of U by man. *Health Phys* 17:619–621
120. Berlin M, Rudale B (1979) Uranium. In: Friberg L, Norberg GF, Vouk VB (eds) Handbook on the toxicology of metals. Elsevier/North Holland Biomedical, Amsterdam, pp 647–658
121. Leggett RW (1989) The behavior and chemical toxicity of U in the kidney: a reassessment. *Health Phys* 57:365–383
122. Domingo JL (2001) Reproductive and developmental toxicity of natural and depleted uranium: a review. *Reprod Toxicol* 15:603–609
123. Gilman AP, Villeneuve DC, Secours VE, Yagminas AP, Tracy BL, Quinn JM, Valli VE, Moss MA (1998) Uranyl nitrate: 91-day toxicity studies in the New Zealand white rabbit. *Toxicol Sci* 41:129–137
124. Gilman AP, Villeneuve DC, Secours VE, Yagminas AP, Tracy BL, Quinn JM, Valli VE, Willes RJ, Moss MA (1998) Uranyl nitrate: 28-day and 91-day toxicity studies in the Sprague-Dawley rat. *Toxicol Sci* 41:117–128
125. Gilman AP, Moss MA, Villeneuve DC, Secours VE, Yagminas AP, Tracy BL, Quinn JM, Long G, Valli VE (1998) Uranyl nitrate: 91-day exposure and recovery studies in the male New Zealand White rabbit. *Toxicol Sci* 41:138–151
126. Paternain JL, Domingo JL, Ortega A, Llobet JM (1989) The effects of uranium on reproduction, gestation and postnatal survival in mice. *Ecotoxicol Environ Saf* 17:291–296
127. Domingo JL, Paternain JL, Llobet JM, Corbella J (1989) The developmental toxicity of uranium in mice. *Toxicol* 55: 143–152
128. Maynard EA, Hodge HC (1949) Studies of the toxicity of various uranium compounds when fed to experimental animals. In: Voegtlin C, Hodge HC (eds) 1953 Pharmacology and toxicology of uranium compounds chronic inhalation and other studies. McGraw-Hill, New York, pp 309–376
129. Maynard EA, Downs WL, Hodge HC (1949) Oral toxicity of uranium compounds. In: Voegtlin C, Hodge HC (eds) The Pharmacology and toxicology of uranium compounds. McGraw-Hill, New York, pp 1221–1369
130. Domingo JL, Ortega A, Paternain JL, Corbella J (1989) Evaluation of the perinatal and postnatal effects of uranium in mice upon oral administration. *Arch Environ Health* 44:395–398
131. Harvey RB, Kubera LF, Phillips TD, Heidelbaugh ND (1986) Validation of impaired renal function chick model with uranyl nitrate. *Bull Environ Contam Toxicol* 36:67–72
132. Morrow P, Gelein R, Beiter H, Scott J, Picano J, Yuile C (1982) Inhalation and intravenous studies of $\text{UF}_6/\text{UO}_2\text{F}_2$ in dogs. *Health Phys* 43:859–873
133. World Health Organization (2004) Uranium in Drinking-water: background document for development of WHO guidelines

- for drinking-water quality. World Health Organization, Geneva, Switzerland, WHO/SDE/WSH/03.04/118
134. Health Canada (1999) Uranium guidelines for canadian drinking water quality: supporting documentation. <http://www.hc-sc.gc.ca/ewh-semt/pubs/water-eau/uranium/guideline-recommandation-eng.php>
 135. Hursh JB, Spoor NL (1973) Data on man. In: Hodge HC, Stannard JN, Hursh JB (eds) Handbook of experimental pharmacology: uranium, plutonium, transplutonic elements, vol 36. Springer, Berlin, pp 197–239
 136. International Commission on radiological protection (1959) Permissible dose for internal radiation. International Commission on radiological protection Publication 2. Pergamon, Oxford
 137. SuLu Fu-Yao Zhao (1990) Nephrotoxic limit and annual limit on intake for natural U. Health Phys 58:619–623
 138. Sample BE, Suter GW, II (1996) Estimating exposure of terrestrial wildlife to contaminants. Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, 59 pp. (ES/ER/TM-125)
 139. Linsalata P (1994) Uranium and thorium decay series radionuclides in human and animal foodchains – a review. J Environ Qual 23:633–642
 140. Haseltine SD, Sileo L (1983) Response of American black ducks to dietary uranium: a proposed substitute for lead shot. J Wildl Manage 47:1124–1129
 141. Harvey RB, Kubera LF, Lovering SL, Mollenhauer HH, Phillips TD (1986) Acute toxicity of uranyl nitrate to growing chicks: a pathophysiologic study. Bull Environ Contam Toxicol 37:907–915
 142. Kovalsky VV, Vorotnitskaya IE, Lekarev VS (1967) Biogeochemical food chains of uranium in aquatic and terrestrial organisms. In: Aberg B, Hungate FP (eds) Radioecological concentration processes. Pergamon, London, pp 329–332
 143. Cooley HM, Evans RE, Klaverkamp JF (2000) Toxicology of dietary uranium in lake whitefish (*Coregonus clupeaformis*). Aquat Toxicol 48:495–515
 144. Cooley HM, Klaverkamp JF (2000) Accumulation and distribution of dietary uranium in lake whitefish (*Coregonus clupeaformis*). Aquat Toxicol 48:477–494
 145. Bywater JF, Banaczowski R, Bailey M (1991) Sensitivity to uranium of six species of tropical freshwater fishes and four species of cladocerans from northern Australia. Environ Toxicol Chem 10:1449–1458
 146. Simon O, Garnier-Laplace J (2005) Laboratory and field assessment of uranium trophic transfer efficiency in the crayfish *Orconectes limosus* fed the bivalve *C. fluminea*. Aquat Toxicol 74:372–383
 147. Vizon SciTec (2004) Final Report on the toxicity investigation of uranium to aquatic organisms. Prepared for the Canadian Nuclear Safety Commission, Ottawa, Ontario by Vizon SciTec, Vancouver, British Columbia RSR-019-80
 148. Tarzwell CM, Henderson C (1960) Toxicity of less common metals to fish. Ind Waste 5:52–67
 149. Parkhurst BR, Elder RG, Meyer JS, Sanchez DA, Pennak RW, Waller WT (1984) An environmental hazard evaluation of uranium in a Rocky Mountain stream. Environ Toxicol Chem 3:113–124
 150. Davies PH (1980) Acute toxicity of uranium to brook trout (*Salvelinus fontinalis*) and rainbow trout (*Salmo gairdneri*) in soft water. Fort Collins, Colorado, Colorado Division of Wildlife F-33-R-15
 151. Liber K, Stoughton S, Janz D (2003) Uranium toxicity testing using early life stage lake trout (*Salvelinus namaycush*). Draft report submitted to Saskatchewan Environment, Saskatoon, Saskatchewan by Toxicology Centre, University of Saskatchewan, Saskatoon, Saskatchewan
 152. Trapp KE (1986) Acute toxicity of uranium to waterfleas (*Daphnia pulex*) and bluegill (*Lepomis macrochirus*). Environmental & Chemical Sciences, Inc., Aiken, South Carolina (Report ECS-SR-30)
 153. Labrot F, Narbonne JF, Ville P, Saint Denis M, Ribera D (1999) Acute toxicity, toxicokinetics, and tissue target of lead and uranium in the clam *Corbicula fluminea* and the worm *Eisenia fetida*: comparison with the fish *Brachydanio rerio*. Arch Environ Contam Toxicol 36:167–178
 154. Hamilton SJ (1995) Hazard assessment of inorganics to three endangered fish in the Green River, Utah. Ecotoxicol Environ Saf 30:134–142
 155. Hamilton SJ, Buhl KJ (1997) Hazard evaluation of inorganics, singly and in mixtures, to flannelmouth sucker *Catostomus latipinnis* in the San Juan River, New Mexico. Ecotoxicol Environ Saf 38:296–308
 156. Holdway DA (1992) Uranium toxicity to two species of Australian tropical fish. Sci Tot Environ 125:137–158
 157. Poston TM, Hanf RW Jr, Simmons MA (1984) Toxicity of uranium to *Daphnia magna*. Water Air Soil Pollut 22: 289–298
 158. Zeman FA, Gilbin R, Alonzo F, Lecomte-Pradines C, Garnier-Laplace J, Aliaume C (2008) Effects of waterborne uranium on survival, growth, reproduction and physiological processes of the freshwater cladoceran *Daphnia magna*. Aquat Toxicol 86:370–378
 159. Kuhne WW, Caldwell CA, Gould WR, Fresquez PR, Finger S (2002) Effects of depleted uranium on the health and survival of *Ceriodaphnia dubia* and *Hyalella azteca*. Environ Toxicol Chem 21:2198–2203
 160. Liber K, George T (2000) Toxicity of uranium to *Ceriodaphnia dubia* at different water hardnesses representative of northern Saskatchewan conditions. Report for COGEMA Resources Inc., Saskatoon, Saskatchewan. Toxicity Centre, University of Saskatchewan, Saskatoon, Saskatchewan
 161. Pickett JB, Specht WL, Keyes JL (1993) Acute and chronic toxicity of uranium compounds to *Ceriodaphnia dubia*. Westinghouse Savannah River Co., Aiken, South Carolina. Prepared for the Department of Energy under contract DE-AC09-89SR 18035, 57 pp (WSRC-RP-92-995)
 162. Semaan M, Holdway DA, van Dam RA (2001) Comparative sensitivity of three populations of the cladoceran *Moinodaphnia macleayi* to acute and chronic uranium exposure. Environ Toxicol 16:365–376

163. Charles AL, Markich SJ, Stauber JL, De Filippis LF (2002) The effect of water hardness on the toxicity of uranium to a tropical freshwater alga (*Chlorella* sp.). *Aquat Toxicol* 60:61–73
164. Franklin NM, Stauber JL, Markich SJ, Lim RP (2000) pH-dependent toxicity of copper and uranium to a tropical freshwater alga (*Chlorella* sp.). *Aquat Toxicol* 48:275–289
165. Hogan A, van Dam R, Masrkich S, Camerilleri C (2005) Chronic toxicity of uranium to a tropical green alga (*Chlorella* sp.) in natural waters and the influence of dissolved organic carbon. *Aquat Toxicol* 75:343–353
166. van Dam R, Hogan A, Houston M, Noul S, Lee N (2006) Chronic toxicity of uranium to *Lemna aequinoctialis* and *Amerianna cuming*. In: Jones DR, Evans KG, Webb A (eds) *Eriss research summary 2005–2006*, Supervising Scientist Report 193, Supervising Scientist, Darwin NT
167. Borgmann U, Couillard Y, Doyle P, Dixon DG (2005) Toxicity of sixty-three metals and metalloids to *Hyalella Azteca* at two levels of water hardness. *Environ Toxicol Chem* 24:641–652
168. Alves LC, Borgmann U, Dixon DG (2008) Water-sediment interactions for *Hyalella azteca* exposed to uranium-spiked sediment. *Aquat Toxic* 87:187–199
169. Burnett C, Liber K (2006) Can *Chironomus tentans* develop tolerance to uranium exposures over several generations? Society of environmental Toxicology and Chemistry (SETAC) North America 27th Annual meeting, Montreal
170. Dias V, Ksas B, Camilleri V, Bonzom JM (2005) Sublethal effects of sediment-associated uranium on *Chironomus riparius* (Diptera: Chironomidae) larvae. *Radioprotection Suppl* 40:S191–S197
171. Hyne RV, Rippon GD, Ellender G (1992) pH-dependent uranium toxicity to freshwater hydra. *Sci Tot Environ* 125:159–173
172. Liber K, Stoughton S, Rosaasen A (2004) Chronic uranium toxicity to white sucker fry (*Catostomus commersoni*). *Bull Environ Contam Toxicol* 73:1065–1071
173. Keklak MM, Newman MC, Mulvey M (1994) Enhanced uranium tolerance of an exposed population of the eastern Mosquitofish (*Gambusia holbrooki* Girard 1859). *Arch Environ Contam Toxicol* 27:20–24
174. Krueger S, Olson GJ, Johnsonbaugh D, Beveridge TJ (1993) Characterization of binding of gallium, platinum, and uranium to *Pseudomonas fluorescences* by small-angle x-ray scattering and transmission electron microscopy. *Appl Environ Microbiol* 59:4056–4064
175. Khangarot BS (1991) Toxicity of metals to a freshwater tubificid worm, *Tubifex tubifex* (Muller). *Bull Environ Contam Toxicol* 46:906–912
176. Liu H-H, Wu JT (1993) Uptake and recovery of americium and uranium by *Anacystis* biomass. *J Environ Sci Health part A Environ Sci Engin* 28:491–504
177. Nakajima A, Horikoshi T, Sakaguchi T (1979) Ion effects on the uptake of uranium by *Chlorella regularis*. *Agric Biol Chem* 43:625–629
178. Horikoshi T, Nakajima A, Sakaguchi T (1979) Uptake of uranium by *Chlorella* cells grown under autotrophic, heterotrophic and mixotrophic culture conditions. *Agric Biol Chem* 45:781–783
179. Chassard-Bouchard C (1983) Cellular and subcellular localization of uranium in the crab *Carcinus maenas*: a microanalytical study. *Mar Pollut Bull* 14:133–136
180. Jaagumagi R, Bedard D (2001) Bancroft area mines (Madawaska, Bicroft and Dyno Mines): assessment of impacts on water, sediment and biota from historical uranium mining activities. Final Report Prepared for Peterborough District Eastern Region Ministry of the Environment
181. Carroll JJ, Ellis SJ, Oliver WS (1979) Influence s of total hardness constituents on the acute toxicity of cadmium to brook trout (*Salvelinus fontinalis*). *Bull Environ Contam Toxicol* 22:575–581
182. Welsh PG, Lipton J, Chapman GA, Podrarsky TL (2000) Relative importance of calcium and magnesium in hardness-based modification of copper toxicity. *Environ Toxicol Chem* 19:1624–1631
183. Holdway DA (1992) Uranium mining in relation to toxicological impacts on inland waters. *Ecotoxicology* 1:75–88
184. Australian and New Zealand guidelines for fresh and marine water quality: Vol 2 – Aquatic ecosystems – Rationale and background information, Section 8.3.7 (http://www.mincos.gov.au/_data/assets/pdf_file/0020/316136/gfmwq-guidelines-vol2-8-3b.pdf). Accessed 14 Jan 2010
185. BEAK International incorporated (1998) Toxicity of uranium and trace metal discharged to the aquatic environment. Prepared for Atomic Energy Control Board, Ottawa, Ontario, August 1998. Dorval, Quebec (Solicitation No. 87055-7-5010/1)

Books and Reviews

- Voegtlin C, Hodge HC (eds) (1953) *Pharmacology and toxicology of uranium compounds chronic inhalation and other studies*. McGraw-Hill, New York

Urban Air Quality: Meteorological Processes

DAVID CARRUTHERS¹, SILVANA DI SABATINO², JULIAN HUNT³

¹Cambridge Environmental Research Consultants, Cambridge, UK

²Dipartimento di Scienza dei Materiali, Università del Salento, Lecce, Italy

³University College London, London, UK

Article Outline

Glossary

Definition of the Subject

Introduction

Characteristic Regions of the Flow and Drivers

Urban Data

Computational Models for Meteorology and Air Flow
in Urban Areas

Examples of Output of Modelling Calculations

Future Directions

Notation and Abbreviations

Subscripts

Abbreviations

Bibliography

Glossary

Mesoscale Scale of weather systems smaller than synoptic scale but larger than microscale or urban scale; tens to hundreds of kilometers.

Neighborhood scale Scale typical of groups of buildings or streets; hundreds of meters to a few kilometers.

Building and street scales Scales of buildings or streets; tens to hundreds of meters.

Fully computational model (FCM) A model which explicitly represents flow and turbulence around buildings.

Fast approximate model (FAM) A model which uses approximations and parameterizations of the fine scale flow to speed up model run times and reduce complexity.

Porosity The volume fraction of air between buildings and therefore a measure of building density.

Urban air quality A general term representing concentrations of pollutants in an urban area. Good air quality corresponds to low concentrations of pollutants.

Urban meteorology Meteorology within an urban area; the urban environment significantly affects mean flow turbulence and temperature.

Definition of the Subject

Meteorological processes in urban areas that are relevant to urban air quality. Of most significance are the impacts of the urban morphology on the mean flow and turbulence, which determines the transport and dispersion of pollutants and therefore their concentration.

Introduction

Concentrations of pollutants within an urban area depend on a number of different factors. These include

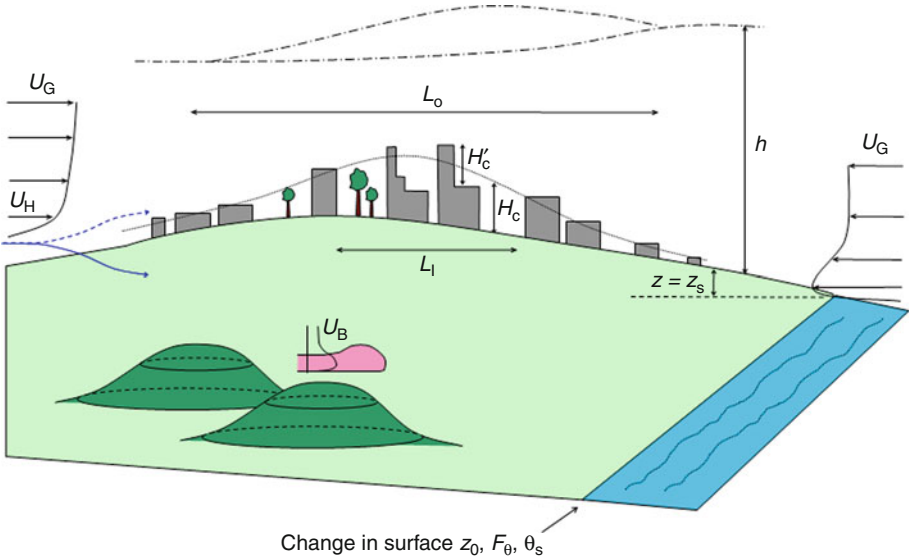
the emissions of pollutants within the urban area, pollutant concentrations transported into the area, and the meteorology within the urban area, in particular, the mean airflow and turbulence (which determines the movement and mixing of the emitted pollutants) and the temperature and solar insolation (which impacts on chemical transformation taking place). This entry discusses meteorology within the urban area with the focus being on the mean flow and turbulence, as these may be substantially different from the upstream flow. Discussed here are the current understanding, experimental and numerical simulation data, and the relative merits of different modelling approaches. For discussions of dispersion within urban areas, see, for example, [1].

Section “[Characteristic Regions of the Flow and Drivers](#)” presents the underlying concepts and prediction methods related to the airflow and meteorology of urban areas that are particularly relevant for understanding and modelling dispersion from the urban sources. It is noted how spatial and temporal patterns of flow and dispersion in urban areas greatly depend on their location (e.g., flat, hilly terrain, coastal, or desert), the types of surface, prevailing meteorological conditions (e.g., strong/weak geostrophic winds), and on the patterns and forms of urban building/street/open space (e.g., isolated buildings or canyons; high or low range of building heights, etc.). It will be convenient to describe the flow in terms of three characteristic ranges of length scale, namely, mesoscale, neighborhood scale, and building/street scales, each with their own general features and where different modelling approaches are necessary, depending on the requirements for accuracy and speed of computation. Section “[Urban Data](#)” presents some details of data, and sections “[Computational Models for Meteorology and Air Flow in Urban Areas](#)” and “[Examples of Output of Modelling Calculations](#)” provide details of modelling methods for flow in urban areas and examples of model calculations.

Characteristic Regions of the Flow and Drivers

Characteristic Regions of the Flow

In urban areas, there are usually considerable variations in the types, sizes, and layout of buildings and streets as illustrated in [Fig. 1](#), and over larger urban areas there are often also significant variations in the natural



Urban Air Quality: Meteorological Processes. Figure 1
Characteristic features of the terrain, buildings, and meteorology on the meso- (regional) scale of significance for urban dispersion. L_0 and L_1 are the overall and inner city length scales, respectively. Note that U_G is the geostrophic wind, U_B is the typical wind speed associated with local buoyancy effects, for example, downslope winds from nearby mountains. H_c is the canopy height and H'_c is the standard deviation of canopy height

Urban Air Quality: Meteorological Processes. Table 1 Scale ranges of urban air flow and dispersion

	Range	Determining factor(s)	Relation to urban and meteorological scales
1.	Mesoscale L_M (~30 km)	Flow over buildings Significant orographic/Coriolis effects Surface influences and pollutants mixed through boundary layer	$L_M \sim L_0$ for large urban areas $l_0 (x \sim L_M) \sim h$
2.	Neighborhood scale L_N (1–10 km)	Flow and dispersion above and between buildings Average geometrical features dominate mean flow/dispersion Growing internal layer above buildings	$L_N < L_0 \lesssim L_M$ Typically $L_N \sim L_1$, inner city scale $l(x \sim L_N) \lesssim h$
3.	Building/street scale L_{BS} (~3–1,000 m)	Flow and dispersion determined by features below building level (except for elevated releases) Mean geometrical and flow features on scale L_{BS} (not smaller) mainly determine dispersion Internal mixing layers below/above buildings are of order of building/street dimensions	$10 H \gtrsim L_{BS} \gtrsim w, b, d$ $l \lesssim H$ (this defines upper scale of L_{BS})

topography. Both these urban scale features and the meteorology have general characteristics over different length scales. These need to be understood and related to each other when developing a picture of the urban airflow as illustrated in [Tables 1](#) and [2](#).

Over level terrain, except in very light wind conditions, the direction and speed of the airflow only changes significantly over the mesoscale (of order 30 km) or greater scales. These changes are determined by Coriolis effects, by the effect of the urban area on

Urban Air Quality: Meteorological Processes. Table 2 Main types and features of urban mesoscale

Increasing effects of orography/heat island, lighter geostrophic wind →				
← Increasing scale	$F = U_G/U_B$ L_O/L_{RO}	>1 (typically windy environment over flat plain)	~ 1 (typically moderate wind over flat surface)	<1 (hilly terrain; near coast)
	$\lesssim 1$	(i) Roughness slow down; boundary layer depth varies.	(iii) Weak flow convergence, change in stability over urban area	(v) Slope flows in open valleys; sea breezes; slope flows and pooling in bounded valleys
	$\gtrsim 1$	(ii) in addition to effects of (i) above anticyclonic turning of wind; weak convergence /divergence on left/right.	(iv) Strong central convergence and cyclonic turning.	(vi) Interactions between orographic and urban heat island flows. Significant variations in orographic and heat island flow across the urban area.

Note: U_G is the geostrophic wind speed,
 $U_B = (gh|\Delta\theta|/\theta)^{1/2}$ is the buoyancy-induced local wind speed
 L_f is the Coriolis advection length ($= U/f$)
 L_{RO} is the Rossby length scale defined by stratification and Coriolis effects ($= hN/f$)

convergence/divergence and turning of the wind, and by temporal variations in the flow forced by diurnal and synoptic effects. Where there are sharp variations in topography the *surface* air flow may change locally (e.g., at the coast or edge of an urban area). However, the whole boundary layer flow does not adjust immediately to the change of surface conditions: this takes place over a mesoscale distance L_M which, in the atmosphere, is typically of order 30–50 km. This distance is usually determined by the Rossby length scale $L_{RO} = hN/f$ where f is the Coriolis parameter and N is the buoyancy frequency of the stable layer at or above the boundary layer of height h . At midlatitudes $f \sim 10^{-4} \text{ s}^{-1}$, and typically $N \sim 10^{-3} \text{ s}^{-1}$, $h \sim 1,000 \text{ m}$. In neutral conditions, L_M is determined dynamically by inertia and turbulence stresses, when $L_M \sim L_f = U/f$, usually a distance of greater than 30 km, while close to large mountains of height H_M , $L_{RO} = H_M N/f$.

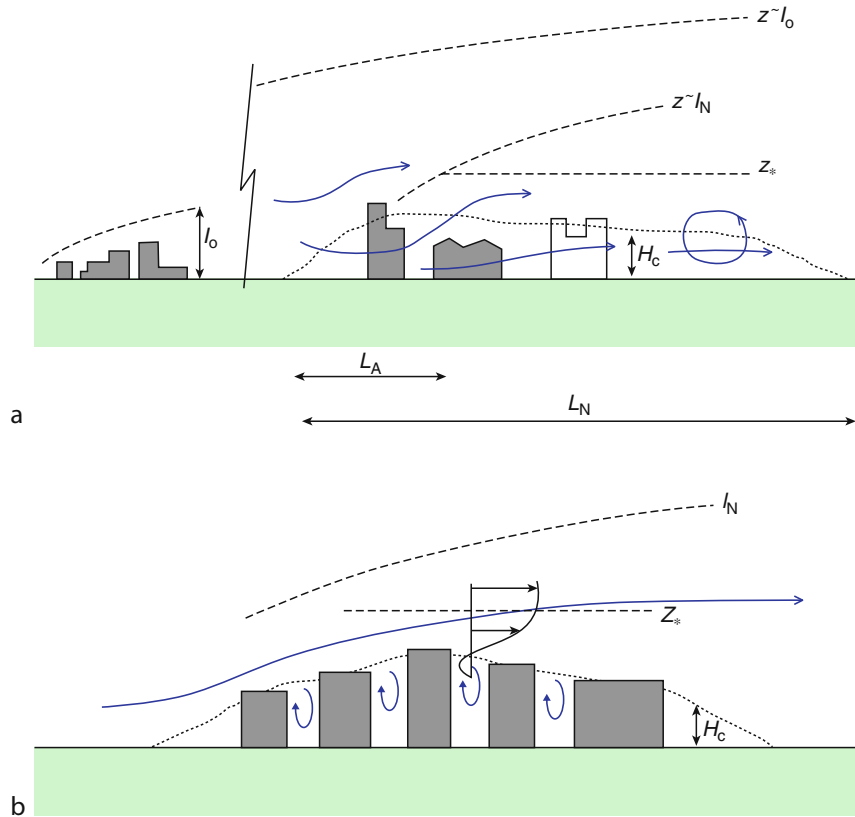
Figure 2 shows that in the zone where the boundary layer is adjusting to the changed surface conditions, there is an internal layer with depth $l(x, y)$ above the ground, separating the air flow and pollutants above and below it. The depth of this layer tends to increase slowly over the area where the surface conditions (temperature or surface roughness) have changed; see,

for example, [2] and [3]. Over the mesoscale L_M , the internal layer reaches the top of the boundary layer, and then the vertical profiles of velocity and temperature over the whole boundary layer are changed.

Over shorter horizontal scales, the internal layer depth is much less than h . Here the perturbed flow has separate characteristics over the intermediate neighborhood scales $L_N (\lesssim 5 \text{ km})$ and over the shorter building/street scales $L_{BS} (\lesssim 300 \text{ m})$. The former scale L_N extends over groups of buildings/streets of similar size and type, and the internal layer depth l_N is greater than the height of the envelope or canopy of the buildings H_C . Note that there is significant mass, momentum and energy transfer between the flow in and above the canopy on this scale. Even though the neighborhood scale includes many buildings, the types of buildings and their patterns have significant effects on this exchange process and on the flow within the canopy.

Mesoscale

In the following section, the main features of the flow and meteorology in the three characteristic flow regions of urban areas are discussed, beginning with



Urban Air Quality: Meteorological Processes. Figure 2

Characteristic features of building envelope and zones of air flow in neighborhood scale. (Note that the definition of this scale depends partly on the level of detail of the computation.) **(a)** "Porous" canopy – with isolated buildings. l_0 and l_N are the heights of the perturbed layers above the buildings; z^* is the height of the top of the shear layer above the buildings. L_A is the adjustment distance. **(b)** Non-porous canopy, where the mean flow passes over the canopy

those on the mesoscale. These are broadly categorized in Table 2 according to the relative effects of buoyancy to inertial forces produced by orographic or heat island effects on the one hand and the effects of Coriolis forces acting over the overall scale L_O of the urban area on the other. The buoyancy and inertial forces are essentially indicated by the value of the Froude number, defined as the ratio of the geostrophic wind U_G to the average velocity U_B produced by the buoyancy forces acting over the urban area, i.e., $F = U_G/U_B$. U_B is driven by the horizontal temperature and hence density changes $\Delta\theta$, which may be natural or caused by the urban heat island, and is defined by $U_B = \sqrt{g l_0 |\Delta\theta|/\theta}$, where l_0 is the internal layer thickness or the height of the nearby mountains if these dominate the flow and g is gravitational acceleration. Note that $\Delta\theta/\theta$ depends on

a number of features of the urban area, the most important being the increased heat storage due to urban buildings, which increases temperature (urban heat island) and the impact of vegetation inside and outside the urban area, which may reduce temperature. Typically the heat island causes increases in temperature of a few degrees in a small town ([4, 5]) so that $|\Delta\theta|/\theta \sim 0.01$ and $U_B \sim 3$ m/s. Larger urban areas show a much higher urban heat island effect (up to 12 C for cities in arid or semiarid environments, [6]) leading to larger values of U_B . Coriolis effects are significant if L_o is comparable or greater than the mesoscale length scales.

Referring to Table 2, it is seen that for typical small to medium scale urban areas (case (i)), buoyancy and Coriolis effects are negligible, $F > 1$, and $L_f, L_{Ro} > L_O$,

the scale of the urban area. For case (ii), as L_O increases, and if there are no thermal effects, Coriolis effects begin to become significant with turning of the wind. Case (iii) corresponds to the case in which the urban heat island, and hence the buoyancy velocity, increases. This will change the stability and boundary layer depth over urban areas. But if the area is even larger so that $L_f, L_{Ro} < L_O$, as in case (iv), then the heat island and Coriolis effects are larger. In this case there is convergence with the flow toward the urban area turning cyclonically (i.e., anticlockwise in the northern hemisphere), as has been measured, especially at night [3].

Strong buoyancy forces occur in the presence of mountains and sloping terrain, so that $F < 1$, leading to marked diurnal variations in the wind speed and direction, over the urban area and outside it. Such areas are associated with sudden changes in the airflow, internal fronts, and pooling of the air in valleys, all of which greatly influence dispersion of air pollution.

Note that because of release of heat stored in buildings at night and increased mixing due to the high surface roughness, the static stability of the air stream usually changes as it moves into and out of the urban area, typically becoming less and more stable, respectively.

Where there are significant buoyancy forces or orographic effects, the boundary layer profiles over the urban area and downwind may not simply adjust toward the equilibrium states found in neutral, stable, and unstable boundary layers over flat terrain. In fact there are characteristic features of the air flow on these scales, especially near hills, coasts, and urban/rural boundaries, that can significantly affect dispersion, such as blocked flow, unsteady slope flows, gravity currents, and boundary layer jets ([7, 8]).

Neighborhood Scale

The characteristic features of the air flow on the neighborhood scale are dependent on the grouping of the buildings/streets (with average building height H , breadth b , width w , and distance between buildings d). In suburban areas or spaciouly designed urban centers, the buildings are effectively isolated. However, in dense suburbs and inner city areas there may be “canyons” with long rows of buildings neighboring streets.

The suburban areas or spacious urban centers are essentially porous to the oncoming boundary layer; the airflow slows down as it passes between the buildings, as a result of both the bulk displacement of the flow over and around the “envelope” or canopy of the buildings and their drag. Downwind of a characteristic adjustment distance L_A (which is of the order of $H(d^2/wb)$ or H/β where $1-\beta$ is the volume occupied by the buildings and β is the volume of air between buildings and therefore a measure of porosity), or typically 10–30 building widths [9], the drag force dominates but is weakened by the sheltering effect of upwind buildings (see Table 3). Over this adjustment distance the air flow above the buildings first accelerates as a result of the vertical displacement by the buildings, and then decelerates as slow moving fluid is expelled from below the level of the building envelope or canopy. Downwind of the adjustment distance (Fig. 2a), if the neighborhood scale is large enough (i.e., $L_N > L_A \sim 30 H$), the air flow (speed U_c) within the canopy is largely driven by turbulent shear stresses generated in the shear layer just above buildings $H_c < Z < Z_* < l_N(x)$ (rather than by the oncoming airflow). Above this shear layer, $Z_* > Z > l_N(x)$, the airflow is characteristic of a surface layer with displacement height z_d so that

$$U(z) \cong \frac{u_*}{\kappa} \ln \left(\frac{z - z_d}{z_0} \right), \quad (1)$$

where z_0 is the aerodynamic roughness length, and κ is von Karman’s constant as discussed by Grimmond and Oke [10], Britter and Hanna [11], and Belcher et al. [9]. Application of Eq. 1 requires knowledge of z_0 , z_d , and u_* . The estimation of these parameters is not straightforward in urban areas. The determination of the friction velocity is also difficult because of the large variability of momentum fluxes inside the urban canopy [12].

The aerodynamic roughness length z_0 is a measure of the turbulence-generating capacity of the ground. The larger the roughness length, the more turbulent the air for a given wind speed. Generally, tall features such as buildings have a larger roughness length than short features. However a densely built area may have a smaller roughness length than a less dense one because the interaction between the air and the region

Urban Air Quality: Meteorological Processes. Table 3 Typical street/building configurations in urban areas

Increasing separation, $d \rightarrow$			
Increasing width w $\downarrow$	b/s	≥ 2	≤ 1
	b/w		
	≥ 10	(i) Canyon ($H/w > 2$) Residential street ($H/w \leq 1$)	(iv) City squares*, parksides
	1–5	(ii) Long blocks (e.g. industrial estates)	(v) Enclosed spaces (e.g. courtyards)
	≤ 2	(iii) Closely-packed industrial sites; ancient settlements**	(vi) Square blocks (e.g. houses), $H/w \leq 1$ Office blocks, $H/w \geq 1$

* Long buildings with large space are rare except for city squares/courtyards.

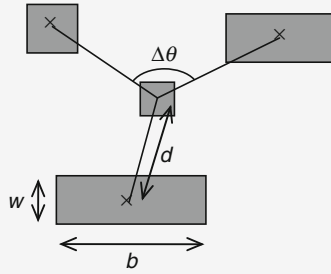
** Wide buildings close together are rare except on certain industrial estates.

In the above s is the distance to the nearest building.

Note:

(a) Mixed cases occur, such as office blocks as part of streets (so H/w varies).

(b) Long and square blocks are usually aligned in streets (i.e. angle between two nearest cases $\Delta\theta$ is 180°).



beneath is reduced. The wind speed close to a rough surface is reduced by friction and a velocity profile can be plotted from the surface to the free fluid. This reduction in wind speed as the surface is approached can be equated to a transfer of momentum from the bulk fluid to the surface, caused by the friction force. The momentum transfer (transport) is large if the roughness is large.

The displacement height z_d according to Jackson [13] can be interpreted as the level of mean momentum absorption. According to the Jackson's formulation this can be determined by the following relation:

$$(z_d - z_s)(\overline{u'w'_s}) \int_{z_s}^{z_d} (\overline{u'w'_s} - \overline{u'w'(z)}) dz, \quad (2)$$

where z_s is the average elevation of the surface and $\overline{u'w'_s}$ is the peak of the shear stress profile. Accurate

knowledge of the aerodynamic characteristics of cities is vital to describe, model, and forecast the behavior of urban winds, turbulence, and the dispersion of pollutants at all scales. The calculation of z_d (which is of order H) based on Eq. 2 is rarely feasible due to the lack of measurements. The classical way to estimate z_0 and z_s in open flat terrain is based on the measurements of wind speed profiles from a tall mast or, less accurately, on the inference from published aerodynamic roughness values for similar terrain elsewhere [14, 15]. Both methods, however, are very difficult to apply to urban areas. A promising alternative that has become available in recent years, due to increased computing resources and the availability of high-resolution 3-D building databases, is based on the calculation of z_0 and z_d from the analysis and measurement of the city geometry (urban morphometry). Morphometric methods express the cities' aerodynamic characteristics in

terms of average building height (H), planar area index (λ_p), frontal area index (λ_f), and other measurable parameters related to the urban morphology (e.g., [10, 16, 17]. While λ_p for a given neighborhood is independent of wind direction, λ_f represents the total area of buildings projected into the plane normal to the incoming wind direction and is a function of orientation. For a given wind direction, λ_f is smaller if the wind angle is oblique, rather than perpendicular, to the front face of the building.

Starting from the lambda parameters z_d and the aerodynamic roughness length, z_0 can be computed using the equations derived by Macdonald et al. [18]:

$$\frac{z_d}{H} = 1 + (\lambda_p - 1)\alpha^{-\lambda_p}, \quad (3)$$

$$\frac{z_0}{H} = \left[1 - \frac{z_d}{H}\right] \exp\left\{-\left[\frac{0.5\beta C_D \lambda_f}{\kappa^2} \left(1 - \frac{z_d}{H}\right)\right]^{-0.5}\right\}, \quad (4)$$

where $\alpha = 4.43$, $\beta = 1.0$, $\kappa = 0.4$, $C_D \sim 1$.

Even though Eqs. (3) and (4) provide good operational estimations for z_{db} as far as z_0 is concerned, there is a warning due to the intrinsic difficulty linked to limited fetch, typical in urban areas, which prevents the flow from being in equilibrium with the changing surface as noted above. Equations (3) and (4) provide a first estimate of surface characteristics, but aerodynamic properties need to be evaluated on a case by case basis. Besides, while z_0 is an important parameter for the above canopy flow description as expressed by Eq. (1), within the urban canopy its use can be replaced by a combination of lambda parameters and their variation with height [19].

Going back to the discussion on the flow characteristics, it is worth mentioning that despite the unevenness and inhomogeneity of these boundary layer flows, the ratios of the r.m.s. values of the three components of turbulent velocity ($\sigma_u, \sigma_v, \sigma_w$) to the friction velocity u_* are quite comparable with their values over level terrain, i.e., $\sigma_u/u_* \cong 2.5$, $\sigma_v/u_* \cong 2.0$, $\sigma_w/u_* \cong 1.3$ [20].

Downwind of the “neighborhood” or urban area where the buildings decrease in height, the mean air flow descends and accelerates. Typically in neutral

conditions the mean velocity adjusts to within 10% of its ultimate (rural) value within about 30 lengths [21].

When the urban area is more densely packed with buildings and they are distributed in the “canyon” form, the urban area is effectively “non-porous” (Fig. 2b). An example of this case is the central area of Nantes [22]. The air does not flow continuously between the buildings but the mean streamlines and the cloud/plume pass above the buildings and air flow is largely above the canopy envelope (average height H_C); the ratio U_c/U_H is much less than σ_w/U_H .

Within the canopy the flow depends on the particular street and building configuration at that scale. There are mean flows along streets at an angle to the mean flows above the buildings, which results in extra lateral diffusion.

The flow above the canopy is essentially equivalent to airflow over a wide hill with length L_N and height H_C , but with a significant value of the roughness length z_0 that is of the order of the thickness of the shear layer over the buildings and the “canyons” between them [23]. The flows in the inner layer above the buildings and in the wake downwind of the neighborhood region are similar to those for the porous case.

Building and Street Scale

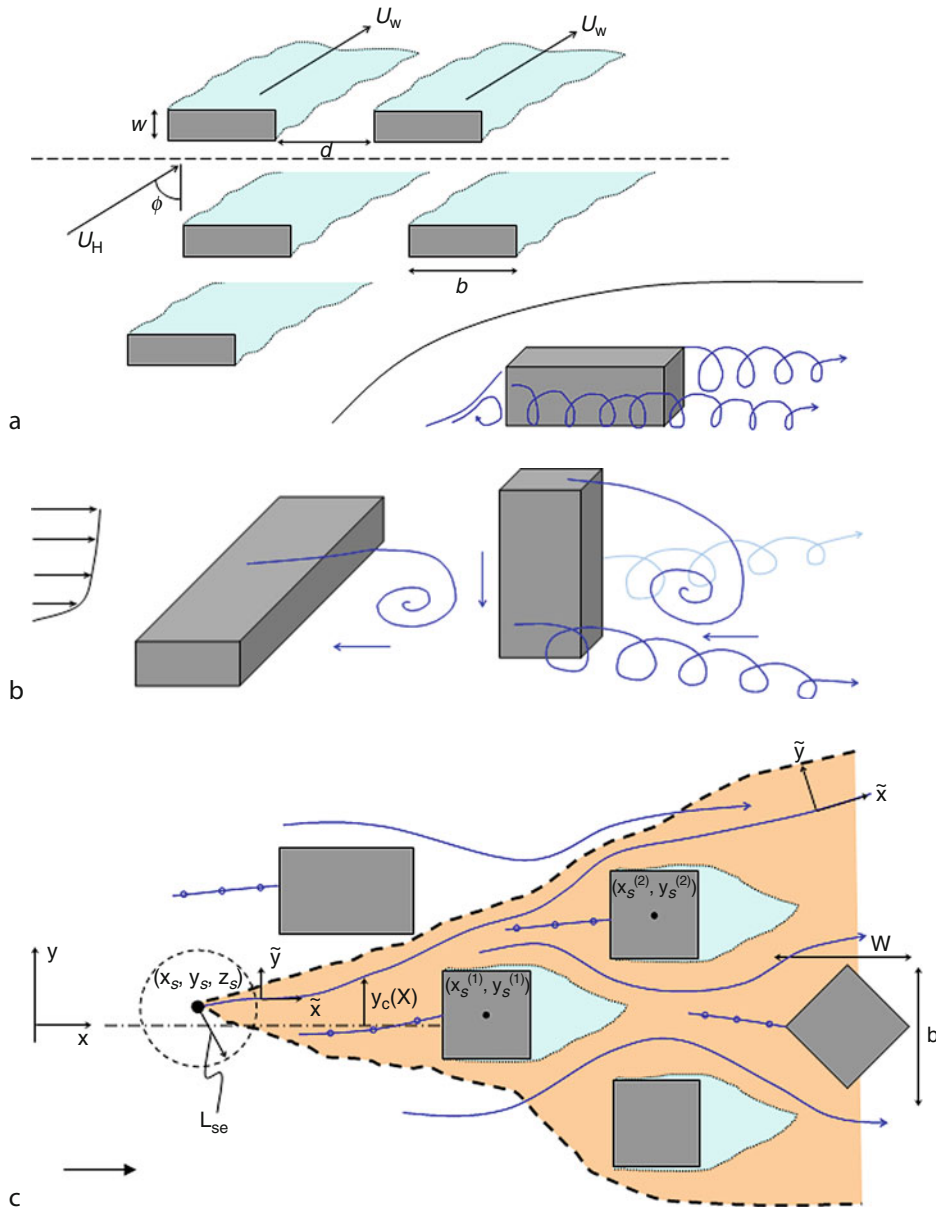
On the building/street scale there are also characteristic features of the flow corresponding to different categories of building street shape and configuration. Various planning criteria and concepts have been proposed for defining these categories, for example, rugosity (or mean canopy height H_C), relative rugosity (defined by building height variability or canopy height variability H'_C), sinuosity (of canyons), the Sky View Factor (SVF), which signifies the fraction of sky dome visible from a specific outdoor position and is important for estimating the amount of incoming solar radiation for energy calculation, and the “compactness index,” which is defined as the ratio of building surface area (excluding the plan area) to the surface area of a cube that has the same volume as the building. For micro-scale phenomena the urban morphometry is the more important, but at the mesoscale both the geometry and surface thermal characteristics play an equal role [24]. These, and related concepts, have been used to guide the suggested categorization shown in Table 3 for

building/street configurations and the consequences for air flow features illustrated in Fig. 3.

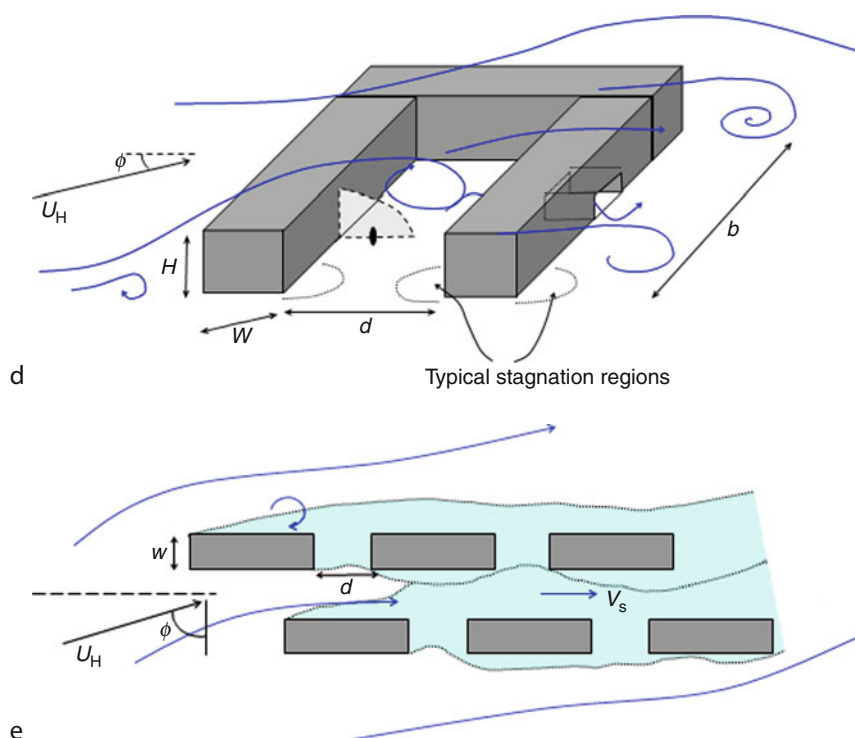
Fluid mechanical studies have shown how flows around individual buildings become significantly distorted in the presence of other buildings depending on the ratio b/d of the breadth b to separation distance d from the nearest building, on the ratio b/w of breadth

to the width w of the building, and on the relative height to width ratio H/w . When there are many buildings, as in urban areas, these flow interactions build up into characteristic flow patterns; this is now examined.

Separated Buildings $b/w, w/b < 2$. The separation distances d are large enough that $b/d \leq 1/3$ ((vi) of Table 3 and Fig. 3a). The flow around each building



Urban Air Quality: Meteorological Processes. Figure 3
(Continued)



Urban Air Quality: Meteorological Processes. Figure 3

Typical building/street configuration. (a) Normal approach angle ϕ small enough (i.e., $\cos\phi > 2w/d$) for independent wakes to be formed (approximately parallel to the approach wind). Note the complex forms of building wakes, in which swirl may persist. (b) Effects of uneven building heights. Note strong reverse and downdraft winds (e.g., [26, 27]). (c) Staggered tall buildings (comparable height), located at $\{x_B^{(i)}, y_B^{(i)}\}$, close to each other. Note how wakes can “disappear” through converging streamlines, while lateral diffusion is enhanced by diverging streamlines (“topological” diffusion). Streamlines are indicated by solid arrowed lines. The streamline through the source at (x_s, y_s, z_s) is $y_c(x)$; $\tilde{x}, \tilde{y}$ are coordinates parallel and perpendicular to this streamline. Streamlines marked with small circles are mean streamlines ψ_{stag} passing through stagnation points where plumes split. The cloud/plume resulting from the source is the shaded area enclosed by heavy dashed line. L_{se} is the effective size of the source outside which atmospheric turbulence determines the dispersion. (d) Enclosed spaces – Typical instantaneous streamlines of the large-scale flow in a courtyard, where $b/d \sim 1$, $b/w \sim 3$, $H/d \sim 1/3$, and $H/w \sim 1$. The shaded area indicates the region of highest concentration of matter released in the courtyard. (e) Normal approach angle ϕ large enough (i.e., $\cos\phi \lesssim 2w/d$) for wakes to merge and “canyon” flow to develop in streets between the buildings

has approximately the same form as that of an isolated building, with recirculating flow regions, turbulent wakes, and horseshoe vortex structures around the base of the building. In slightly stable conditions the vortex structures can persist far downwind but in most urban cases where the flow is neutral and highly turbulent they are not significant. On the other hand the turbulent wakes from upwind buildings do impact on those downwind, enhancing the

mixing. Isolated buildings, especially in slightly stable conditions can produce swirling wakes far downwind [25]. When there are marked variations in the height of adjacent buildings, the wake vorticity shed from upwind buildings can produce sharp down-flows and increased trailing vorticity in the flow direction as discussed by Lawson [26]. These effects contribute to mixing between the canopy and external flow.

For tall buildings that are not closely packed and where $H/b > 1$ (Fig. 3b), there is strong mixing in the horizontal plane and to a lesser extent in the vertical direction because of the high turbulence generated. Buildings placed sufficiently close to each other may result in wake interaction, which tends to cause downwash [27] and strong swirl around the sides of tall buildings. This enhances vertical mixing in the lower part of the downwind wakes and through lateral convergence reduces their downwind extent. When $b/d < 1/3$, corresponding to long buildings that are sufficiently separated from each other, the flow over an upwind building can descend into the space between the buildings. This flow is significantly sheltered and stagnation areas are larger than for isolated buildings.

Close Packed Buildings not Aligned. In some city centers and some types of industrial plant, the spacing between buildings are within about two or three widths (i.e., $b/d \gtrsim 1/3$), but the gaps between the building do not form continuous “canyons” (iii) of Table 3). The air flow passes around the buildings and the streamlines move irregularly between the buildings, converging and diverging as they do so, as sketched in Fig. 3c [28]. The wakes tend to be reduced in length because of cancellation of vorticity.

Where the separation distance is smaller, i.e., $b/d \geq 1$, the flow is further changed because there may be other complicating geometrical factors, such as the effect of a courtyard, which further slows up the flow, or open gateways, which can lead to sufficiently high local winds that affect the overall flow pattern (Fig. 3d). Here the interaction between the vorticity layer from the top of the buildings and impacting on surrounding buildings leads to even more complex recirculating flows than in street canyons. However, because of the finite extent of these flow regions, it can generally be assumed that (unlike elongated street canyons) there is only a small mean velocity component U_C below the building height, i.e., $U_C/U_H \lesssim 0.1$. However, the turbulence level is quite high in these recirculating flows, with $\sigma_u \sim \sigma_v \sim (0.5\text{--}1.0) \cdot \sigma_u(z_*) \sim 1.2\text{--}2.5 u_*$, so that $\sigma_u \sim V_s$ [23].

Flow Aligned with Rows of Buildings ($b/w > 1$, $H/d \gtrsim 1/3$). In this case buildings are located in rows with separation gaps d that are comparable with their height H (i.e., streets). The case includes buildings that are loosely aligned in rows or continuous canyons

((ii), (iii), and (iv) of Table 3). Where the downwind buildings are nearly aligned with the wind (Fig. 3e, $\cos \phi \leq w/d$), they distort the wake’s vorticity so that the wakes tend to be aligned with the building [29]. The recirculating flow in the wake of the upwind row tends to extend across the “street,” and there is usually a significant mean velocity V_s along the street. The mean flow has a component along the street and, when the external flow is not aligned with the “canyon,” a swirling motion in the perpendicular direction that tends to extend up to the top of the buildings. The peak mean and fluctuating velocities near the ground are comparable with those at the level of the top of the buildings. The key quantitative parameter for dispersion modelling is the ratio of the mean wind along the street V_s to the mean wind U_H above the buildings at the height H (in the approach flow). Computation and field observations [23] have suggested a large range of this ratio from 0.3 to 1.0. The analysis also suggests that the value of V_s can be quite sensitive to the value of V_s at the upwind end of the canyon. Thus a major crossroad can have a large effect on the canyon flows downwind.

In conclusion it has been shown that most of the usual types of building/street shape and configuration found in urban areas can be categorized in geometrical terms that also correspond to the characteristic air flow patterns found in these situations for typical wind conditions. In particular there is a major distinction between, on the one hand, flows determined by random interactions of wakes of well-separated buildings and, on the other hand, flows where the interaction of wakes is of less relevance, for example, highly organized “canyon” like flows that occur when the buildings form street canyons or, if they are separated, when they are positioned along streets and flows in enclosed areas (e.g., courtyards) or between very densely packed buildings.

Urban Data

Experiments and dedicated simulations have shown features of urban dispersion at the neighborhood and street buildings scale. This includes how the mean transport direction differs from the mean flow above the buildings, the rapid vertical and horizontal spreading of plumes in wakes and canyons, the more complete mixing by the interacting wakes (with lower relative

levels of fluctuation compared with rural terrain), the transition of dispersion below the buildings level to above them and the variation in concentration downwind of sources and within the buildings.

In the context of urban air quality, the full use of models based on an up-to-date understanding of the flow as described in the previous sections is somehow limited by the scarcity of available data. Urban flow and dispersion models require input information and source distribution data about the mean flow field, turbulence, and source distribution. Until very recently velocity field data in urban areas were quite sparse and typically derived from the experiments in the USA. Some progress has been made in recent years with some field experiments performed in some European cities even though one has to accept that datasets from field experiments are yet not fully complete given the inherent difficulty and the cost of urban flow and dispersion experiments. This aspect has motivated the increase of laboratory experiments of flow and dispersion over simplified groups of buildings (building arrays) as well as over scaled reproduction of real city quarters. In addition to field studies, laboratory experiments allow easier validation of physical or numerical models because of carefully controlled flow conditions and the possibility of making numerous measurements. The laboratory experiments can be either wind-tunnel experiments [30–32] or water-channel experiments [33]. Some field studies have focussed on real cities, releasing a tracer gas in selected neighborhoods, for example: Copenhagen [34]; Salt Lake City, “Urban 2000” [35]; Oklahoma City, “Urban Joint 2003” [36]; Basel (BUBBLE experiment) [37]; London (the Dispersion of Air Pollution and its Penetration into the Local Environment known as the “Dapple” experiments) [38, 39].

The advantage of such experiments is to simulate flow and pollution dispersion in real conditions, but an important drawback is that the complexity of the urban morphometry can obscure the “true” impact of buildings from other complex phenomena due to the specific configuration of the selected city or neighborhoods. In the case of the Mock Urban Setting Test (MUST, [40]) it was determined that the focus would be on a simplified representation of a city by modelling it as a regular array of shipping containers, while keeping meteorological conditions real. This study has

helped to clarify and give scientific evidence of specific flow phenomena as outlined in the previous sections. One important aspect for dispersion application has been the quantification of the deflection of the mean plume axis relative to the incoming wind direction induced. This can be significant, being up to 23° for a wind direction of 27° relative to the axis of the array, as found in the MUST experiment. Extrapolations to the real conditions are not always straightforward. The full field experiments mentioned above confirmed that, depending on building density and morphology, transport and dispersion may lead to complications not observed in simpler building array distributions. Near street level, for example, the dispersing plume may travel several blocks in a direction opposing the prevailing wind and many blocks laterally. This topological dispersion can lead to secondary sources and significantly alter the rate of lateral dispersion. A ground-level source can rise several hundred meters in depth in less than a block when caught in the updraft just downwind of a tall building. Buildings also alter the timing of the transport and dispersion, generally resulting in much longer residence times as compared to open terrain. Even though it is difficult to generalize the time of residence of a pollutant cloud, the Birmingham experiments suggested a few hours for a single source. However, the Dapple experiments did not give a definite conclusion on this. A further observation important for air quality applications is that the experiments have shown the presence of stable stratification during the night when consequently street level flow is less well coupled to the upper level reference. This might be more frequent in US cities than in European cities. As observed by Barlow et al. [39], stable conditions occur in London only occasionally. The analysis by Allwine et al. [35] carried out using data from the Joint Urban campaign in July 2003, Oklahoma City, showed that a stable layer is often observed overnight, with a nocturnal jet. Thus, there is no common pattern regarding the frequency of stable conditions; such conditions are a strong function of thermal advection due to regional scale flow processes and also depend on the specific materials used in the different cities.

The above are only some examples and are far from being a complete discussion of specific results obtained from the urban experiments.

Computational Models for Meteorology and Air Flow in Urban Areas

This section discusses those computational models for both meteorology and air flow in urban areas that are used as input to models for calculating dispersion. The numerical and physical basis and operational attributes are summarized in Table 4 for the three ranges of length scale introduced in section “Characteristic Regions of

the Flow and Drivers”, and for the different types of model, ranging from off-line, fully computational models (FCM) to the fastest approximate models (FAM) that can be run on PCs. The choice of the model depends on their practical purpose (in relation to dispersion modelling), the required level of spatial/temporal detail and scientific understanding involved, and on the detail and accuracy of meteorological and topographical input data. For research and for off-line

Urban Air Quality: Meteorological Processes. Table 4 Types of computational models for urban meteorology and air flow

1. Mesoscale
(i) Full computational models (FCM) Based on grid box numerical methods; full physics (fluid dynamics and thermodynamics). Input: from synoptic (or regional) numerical weather prediction models and local surface conditions, for example, surface fluxes for heat, F_θ, water vapor, F_v, roughness length, $z_0(x, y)$, and surface elevation, $z_s(x, y)$. Output: mean flow, turbulence statistics from above building envelope up to tropopause as a function of space and time over the urban area. Typical number of grid boxes for 1/3–1 km resolution on 30–100 km grid is 10^6–10^7. Takes about 3 h per 1 simulation hour, faster for lower resolution model.
(ii) Fast approximate models (FAM) (a) Flow perturbation models (suitable where $U_G/U_B \gtrsim 1$) using fast/approximate numerical and analytical methods, simple physics (for turbulence, stratification profiles), input from observational data (reduced) or NWP models and from local surface conditions. (b) Local flow/thermal models (especially where $U_G/U_B \lesssim 1$) using formula/fast approximate models for special conditions/physics, for example, slope winds, sea breeze, and thermal convection; often adjusted for local conditions. (c) Mass consistent models to compute flow fields given data at several points in the flow domain and surface elevation z_s. (Only suitable if such data exist for all relevant cases.) (d) Typical resolutions can be on a horizontal scale of 100 m or less. (e) These models can be run over a few minutes on PCs for each case. Current simplified models do not account for unsteady conditions over large urban areas (where $L_O/L_f, L_O/L_{Ro} \gtrsim 1$).
2. Neighborhood scale
(i) FCM using CFD methods Grid point numerical models with approximate representation of buildings and open spaces (as porous medium; as distributed forces; or as approximate shapes); approximate turbulence and heat transfer models; input from mesoscale models or measured data. Output: mean and turbulence profiles within the building envelope and up to the boundary layer/inversion layer height. Typical horizontal resolution is greater than the spacing between buildings. For a “neighborhood” of 5 km (e.g., city center) with 10^6–10^7 points, the resolution would be 100 m for an “accurate” computation.
(ii) FAM (a) Flow perturbation models (applicable where $\ln z_0$ or F_θ vary significantly) using fast/approximate numerical and analytical methods with buildings/open spaces described by average properties and their relative fluctuations (e.g., average building height H, spacing d, and also typical r.m.s fluctuations in H and d); simple turbulence and heat transfer models for canopy flows; input from mesoscale models or local data, and estimates of distributed effects of buildings. Output: mean flow within and above the canopy and turbulence above the canopy. (b) Local flow/thermal models (where $\ln z_0$ or F_θ are approximately uniform). Formulae based on local dynamical and thermodynamic balance and estimates of distributed effects of buildings; input based on average wind, thermal flux, and estimates of distributed effects.

Urban Air Quality: Meteorological Processes. Table 4 (Continued)

3. Building/street scale
(i) FCM using CFD methods Grid box numerical models with accurate representation of buildings, approximate turbulence and heat transfer models; input from neighborhood models or assumed data. (Very fast versions using idealized (e.g., inviscid) equations with finite numerical diffusion and approximate boundary conditions for buildings.) Output: mean flow and turbulence profiles around buildings of different shapes and grouping within an urban area. Typical computational domain is greater than building length or spacing L , or d (100 m) and grid size of 1 m or less. Note: sensitivity to inflow conditions from other buildings and atmospheric turbulence. Models better for clusters of buildings than isolated buildings (where large-scale atmospheric turbulence has less influence). Simpler methods are accurate/fast enough for practical use in off-line diffusion calculations.
(ii) FAM (a) Turbulent wakes of individual buildings based on perturbation methods and typical flows near buildings. (b) Closed packed buildings modelled by ideal (potential) flow. (c) Idealized models for interactions when wakes of upwind building impinge on downwind structures. (d) Canyon models (semiempirical formulae – not yet well established). (e) Canyon/street intersection models (e.g., ideal potential flow).

validation studies, a variety of types of model tend to be used: some studies require great detail and accuracy about particular or commonly occurring situations; others require simpler models for rapidly calculating a wide range of meteorological and topographic boundary conditions. When models are analytically based, their predictions can be easily understood and may be expressed in useful formulae.

The reason why it is necessary to have different types of model for the different distances is because every type of model can only represent a finite range of length scales, limited by the capacity and speed of the computational systems and data input.

Mesoscale Models

On the mesoscale range, the inputs to the models are the representation/parameterization of the topography and the natural features of the surface necessary for computations down to the smallest scale L_M/R_M (which typically is of order 1/3–1 km). This is therefore the scale over which the surface conditions are averaged, including ground surface elevation $z_s(x, y)$, roughness length $z_0(x, y)$, surface heat flux $F_\theta(x, y, t)$, and surface temperature $\theta_s(x, y, t)$ (which is usually derived in operational mesoscale models by coupling the atmospheric model to a thermal model for ground temperature, allowing for radiation from/to the earth's surface). In fully computational models, FCMs, the

boundary conditions at the edges of the domain (and, in some models, above the domain) have to be specified and are usually derived from larger scale regional or global numerical models. These are usually updated at regular intervals, for example, every 3–12 h, with observational data taken at all levels in the atmosphere. This process of “data assimilation” is also beginning to be applied directly to mesoscale models, for example, in urban areas with many measurement sites.

A critical feature of any FCM, especially in the boundary layer and when the flow is influenced by mountains, buildings, surface heating etc., is the representation of the effects on residual (or computed) scales greater than L_M or R_M of the turbulence at the smaller space and timescales. Some models, such as HOTMAC [41], contain quite complex sub-models of the turbulence statistics (with extra equations for the turbulent kinetic energy and turbulent dissipation rate $(k-\epsilon)$), while others, for example, WRF [42], COAMPS [43], MM5 [44], MESONH [45], and UK Met. Office unified models [46], use eddy viscosity or even simpler parameterizations (e.g., assuming a known form of the velocity profile). The experience from field studies in the USA is that for urban areas located on sloping terrain where buoyancy forces are significant (i.e., $F \lesssim 1$), the models with the more complex turbulence models are more accurate in these conditions. But the greatest improvements in accuracy, especially in predicting the mean wind speeds and wind direction

in changing meteorological conditions, and in complex terrain, arise from using these models with finer resolution of the order of 1 km or less. However even with finer resolution and complex modelling, such models still cannot predict some significant features of surface layer turbulence, such as evening transition of upslope flows and formation of gravity currents, or calculate relevant statistics of the turbulence for dispersion models, unless they are used as large eddy simulations [47], which, because of computational requirements, is currently only possible in research mode.

As explained in Table 4, these models take many hours to compute a single meteorological situation. Nevertheless, they are sufficiently reliable indicators to be used operationally. Fast approximate models, FAMs, for air flow and meteorology over mesoscale distances are being developed, based on recent research. These models are useful as qualitative guides to the complex flow that might occur on the mesoscale, for quantitative predictions as input to dispersion computations.

As shown in Table 4, when local buoyancy effects are weak (i.e., $U_G/U_B \geq 1$), the air flow over an urban area is generally a perturbation of the oncoming flow. Then, from semi-analytic models using perturbation methods, faster computational schemes have been devised, and are being developed in a general way to allow for orography, roughness change, and some effects of surface heating [8], [48]. Although these are perturbation methods, the changes in wind speed and direction predicted (and verified) by these models can be quite large ($\sim 50\%$ or more).

The perturbation modelling approach being adopted in this range of meteorological flows is similar to that used for sub-mesoscale/neighborhood scale orographic flows (for example, in the FLOWSTAR model [49] and RIMPUFF [50]), the main difference being that over the larger scale the internal layer l reaches the top of the boundary layer h , so that the inversion height is affected and Coriolis effects have to be included (which, for example, significantly influences flows along coasts and up and down large river valleys, [51]). Typically these models are run in the steady state. Once the time dependence over several hours is significant, a fully computational model is necessary.

When buoyancy forces are significant, $U_G/U_B < 1$, no FAM is yet available to cover all the most significant

types of air flow and meteorological conditions that might be needed for dispersion calculations. However, models and physical estimates have been derived for some particular situations, especially when Coriolis forces play a small role, for example, in wide valleys with low slopes, large heat island effects, and sea/lake breezes [48, 52, 53]. Not only are the quasi-steady features of these flows now quite well described, but also their time/space dependence over the diurnal cycle. In the absence of strong geostrophic winds (or very stable local conditions) the local diurnally varying buoyancy forces may control the flow and therefore they are quite predictable (e.g., a valley wind, or sea/lake breeze on a still day). Since the single most important aspect of the air flow needed for dispersion calculations is the direction of the wind, such physically based models provide vital information for estimating the dispersion in complex terrain.

Neighborhood Scale Models

On the neighborhood scale (typically 5 km), fully computational models, even with the latest computing systems, cannot generally resolve the flow around every building. Typically, for computations with 10^6 – 10^7 grid boxes, the horizontal grid spacing is about 100 m or greater in the horizontal direction. To resolve the velocity field around a building in order to calculate its drag effect on the overall flow, the grid boxes would have to be of the order of 1 m or less. Therefore even FCMs can, at present, only calculate average features of the flow over these grid box scales by estimating the average effects of buildings and streets in the model. This is expected to change in the near future with further increases in computational power.

For most practical applications in urban areas, different modelling approximations are needed (even with an FCM) depending on whether the buildings in the neighborhood being considered form effectively porous or non-porous regions of resistance. In the case of effectively porous buildings (Fig. 3; category (i) of Table 3), the effect of the buildings is estimated by a drag coefficient averaged over the whole volume occupied by the buildings and the space between them (typically this is of the order of the porosity β).

The FCM input is the mesoscale flow field approaching the neighborhood region (or the flow

leaving the adjoining neighborhood), for example, suburbs adjoining a city center as in Fig. 1. FCM output will be the mean flow and basic turbulence statistics that depend on the particular closure model used, for example, mixing length, turbulent energy-dissipation (k - ϵ) closure, or Reynolds stress closure [54, 55]. All three of these models lead to estimates of turbulent kinetic energy, but only the latter also provides estimates of length scale, and the space/time dependent evolution of turbulence structure, which is significant where the areas of building density or height change sharply. The third method allows for the anisotropy of turbulence, which changes in these transition zones. Such models predict the mean velocity field everywhere (spatially averaged over the grid box) and the turbulence outside the canopy. Within the canopy, where large-scale inhomogeneous turbulence is generated by the shear layer over the top and by eddy motions around buildings, the spatially averaged models are deficient.

In the second case, an effectively non-porous neighborhood, the mean air flow passes over the building envelope as if over a hill with elevation $z_s(x, y)$. The roughness length z_0 also changes as a result of turbulence and recirculating flows in the courtyards, streets, etc. between the buildings. Estimating these parameters z_s, z_0 in terms of the building layout is only approximate, (e.g., $z_s \cong H_c, z_0 \sim H/30$). As with porous built-up areas, the parameterizations for calculating average flow properties can be estimated approximately by detailed computation of typical local areas. These calculations can also provide estimates of the local turbulence, needed for dispersion computations [23]. FCMs using turbulence closure models have been extensively applied to the kinds of recirculating flows that occur between buildings in non-porous urban areas; most validation has been for wind tunnel tests and engineering flows, where the kind of very-large-scale eddies and downdrafts found in atmospheric flows are absent. Such motions can lead to more rapid exchange between the upper flow above and within the canopy layer, as discussed in the review by Mestayer and Anquetin [5].

As shown in Table 4, and reviewed by Britter and Hanna [11], most FAMs for the neighborhood scale tend to focus on equilibrium flows within and just above the canopy. In a “porous” canopy (Fig. 2a), the mean velocity within the canopy $U_c(z)$ is driven by

the turbulent shear stresses generated in the intense shear layer just above the canopy. Here the ratio of $U_c/U(z_*)$, where $U(z_*)$ is the velocity at the top of, or above, the roughness layer (Fig. 2), depends on the porosity – a typical value being 0.3. In a non-porous canopy, typically in the inner city, where flows are determined by canyon and downdraft effects, this mean velocity ratio varies considerably over the range 0.1–0.3 [23].

The approximate average velocity profile (Eq. 2.1) is not applicable at the edges of “neighborhoods” where the density and heights of buildings vary, or in the interiors of such regions, such as near parks, squares, etc. At the edges of porous canopies, where the air flow in the canopy varies rapidly, an FAM approximate model has been developed based on the same approach as FCM by representing the canopy as a porous layer and (as for the FAM in the mesoscale range) solving perturbation equations semi-analytically to provide fast computation of the mean velocity and shear stress [9, 56]. These have been verified against field and wind tunnel studies.

Over non-porous canopies (typically categories (i)–(v) in Table 3), the determining parameters ($H_c, z_0(x, y)$) are the same as for FCM over these elevation/roughness changes. The linearized FAM approaches give very similar results to those using FCMs [57, 58].

Modelling the Building and Street Scale

The purpose of detailed flow modelling in and around individual buildings and the surrounding streets is firstly to understand how rapidly released matter disperses locally. Secondly, detailed studies of these local scale flows are necessary, as explained in the previous section, to develop models over neighborhood scales for flow and dispersion.

Fully computational methods, using a variety of turbulence modelling methods, have been extensively applied to computing flows around single buildings in turbulent boundary layers. Using turbulence closure methods, many models have predicted mean velocities near the buildings within one or two widths; but such models (e.g., k - ϵ models) have tended to over-predict small-scale turbulence around the structure [59] and under-predict the large-scale atmospheric eddies. This leads to an over/under-prediction of the mean velocity

defects in the wakes downwind of 3-D/2-D structures in the atmospheric boundary layer. Only the computationally intense method of Large Eddy Simulation (LES), or unsteady modelling of the fluctuating flows using turbulence closure models, can accurately represent the distortion of the large-scale eddy motion around the building and its effect on downwind wakes [60, 61].

Calculations using FCM with turbulence closure have also been performed on flow around small groups of buildings using resolutions down to about $0.1 H$. The predictions of the turbulence in these flows are more reliable than those of isolated structures because most of the turbulence in this instance is generated locally between the buildings and is of smaller scale (similarly, wind tunnel models of dispersion around buildings are generally more reliable when the buildings are in groups than when isolated because wind tunnels cannot simulate the low frequency fluctuation in wind direction). This is because the primary interactions between the wake of one building and its impact on adjacent buildings are not sensitive to the structure of turbulence; however, the wakes of isolated or well-separated buildings are strongly affected by the form of the low frequency spectrum of turbulence. This is why in some institutions, for example, the US Naval Laboratory [62] and more recently at Los Alamos [63], work is being undertaken on dispersion in dense urban areas with fast FCMs in which approximate velocity fields are computed using classical grid box methods of finite size (say 1 m) but neglecting completely any modelling of turbulent stress. This reduces the number and the order of the equations to be solved and speeds up the computation by a factor of 10 or more. The finite size of the grid effectively introduces numerical diffusion that simulates many of the same mean flow features developed using the more complex turbulence closures.

Another simplification to speed up the computation of the mean velocity field is to represent the effects of the individual buildings as a distributed force acting on the flow (or source/sink distribution) and thereby only approximately representing the shape of the building [64]. In most FCM calculations, considerable computational resource and time is spent on exactly representing the building shape, while at the same time making considerable approximations in the flow.

There are at present no general FAMs that are capable of calculating the detailed flow around any isolated building or group of buildings, or even calculating those broad features of the flow needed for modelling dispersion close to the building. The main reason is that the unsteady separated shear layers that control the flow are of much smaller scale than the building. This means that the recirculating flow region, which greatly affects the magnitude and timescale of concentration fluctuation, is not only highly unsteady but also very sensitive to the effects of nearby buildings. Nevertheless, the main features of the mean flow patterns and typical magnitudes of the mean and fluctuating velocity have been classified and partially quantified for many types of buildings and groupings. Therefore, although no practical FAM for the mean flow is available at present, these features of the flow field are now well enough known to derive useful quantitative FAMs for the dispersion. For example, when buildings are far enough apart that their near-flow fields and their wakes are approximately independent, the wakes can be approximated as a perturbation to the approach flow, a theory by Counihan et al., 1974 [21] that is used in the “Buildings” module in ADMS [67]. The same model can be used approximately with linear superposition when buildings are separated by at least $3H$, where H is their height. As the ratio of breadth to height, b/H , increases from 1.0 to 10, this minimum separation distance $d_{\min}$ increases linearly to about 10 H (Fig. 3a(i)).

When d is less than $d_{\min}$ (Fig. 3a(ii)), it is found that the recirculating regions of the wakes of the upwind buildings are elongated and extend to the nearest downwind building [65]. For buildings designed in rows nearly parallel with the wind direction, the flow structure consists of turbulent recirculating regions in the spaces downwind of the buildings and relatively high-speed streams in the open streets between the buildings [28]. Describing the flow structure in this way provides the basis for FAM for dispersion calculations.

By contrast, when the buildings are placed in a staggered pattern relative to the wind direction, the wakes tend to disappear and the mean flow between the buildings consists largely of unidirectional bifurcating patterns of diverging and converging streamlines, with small regions of recirculating flow. Such flow can be

computed by fast potential flow methods [28], which are even faster if the shape of the layout of the buildings is approximated [64].

When the buildings are long enough and close enough (i.e., categories (i), (ii), and (iii) of Table 3) and the wind is at any angle, or if they are situated in rows and the wind is at a small glancing angle to the row (Fig. 3a(ii)), then the characteristic canyon flows are formed in the streets between the buildings. When the buildings either side of the street are of comparable height, a strong feature of the helical canyon flow is that it does not extend above the buildings. Therefore the appropriate FAM for these flows is to assume a canyon flow below the building level and a rough wall boundary layer flow above it [23].

The final consideration is of the air flow in typical courtyards and “squares” connected to adjoining streets (Fig. 3c). Simulations confirm that these are regions where fluid trajectories are well mixed. This provides a basis for constructing FAMs for dispersion in these regions using semi-analytical diffusivity models.

Examples of Output of Modelling Calculations

A series of examples of recent model calculations showing the impact of buildings on flow are now reported. The cases considered are a single building, a regular building array, and a complex urban junction. The examples mainly comprise simulations using an FCM (specifically a CFD model); one example using an FAM is also presented.

CFD models do not describe all meteorological phenomena as described earlier; however, they are useful tools for predicting flow characteristics in street

canyons and within urban areas at the neighborhood, building, and street scale. They can also be used to interpret wind tunnel and field data and their outputs can therefore be used to improve operational models. Most available general purpose CFD models currently still need to be validated against wind tunnel or field data to obtain confidence before applying them to a particular case study [67]. Some FAM models used for calculating dispersion at the building and street scale are reviewed in [68].

Single Building

Table 5 summarizes cases investigated by means of both wind tunnel and CFD simulations by Wang and McNamara [69]. The Launder and Spalding’s standard k - ϵ model [70] for turbulence and the advection-diffusion model for dispersion was used for the computation with a computational grid comprising 500,000 cells and smallest dimensions equal to 0.005 m. The cases comprised a cube (1), a tall building (4), a wide building (5), and intermediate cases (2 and 3).

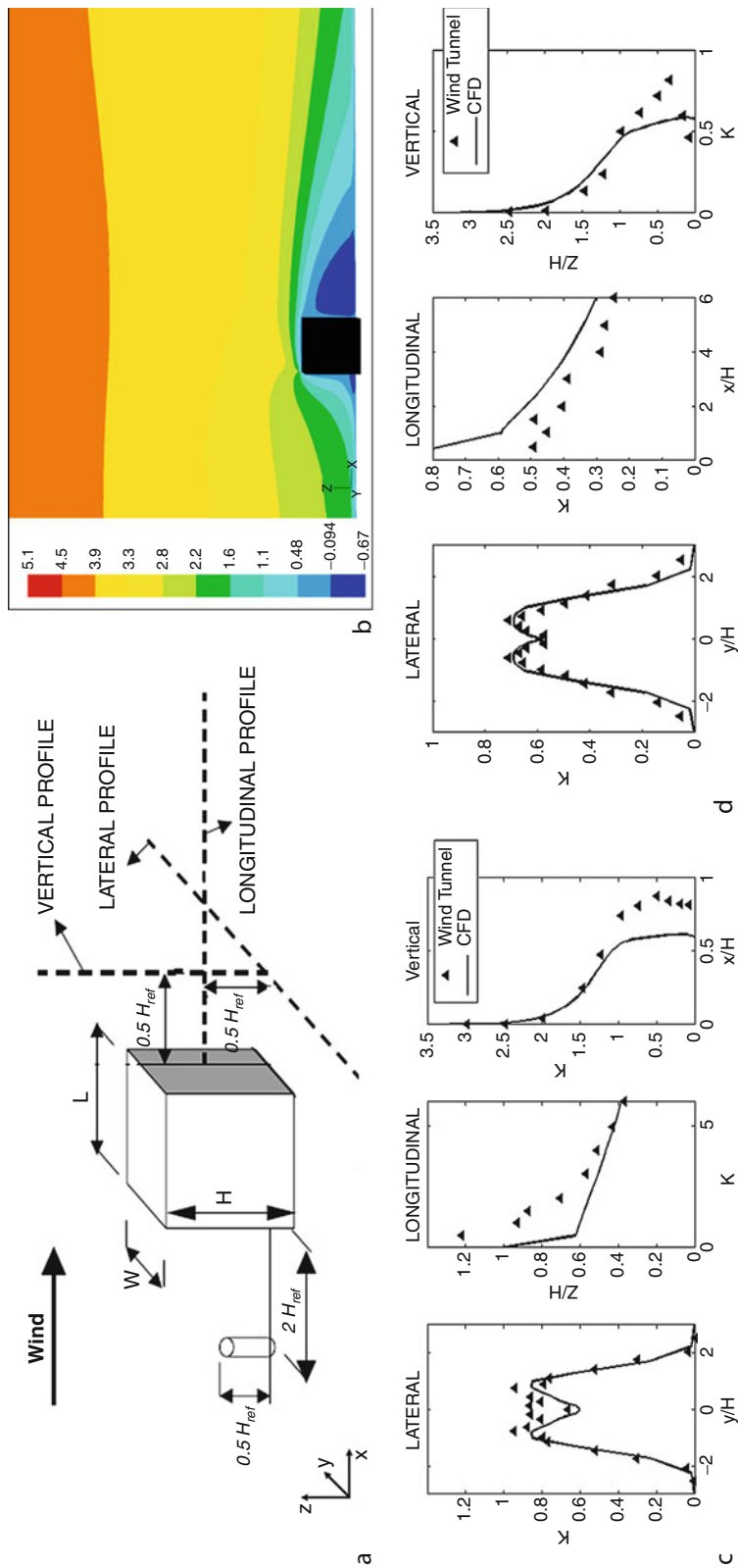
Mean concentrations are expressed as dimensionless values K which are defined as:

$$K = \frac{CU_{ref}H^2}{Q},$$

where Q is the emission rate, C is the measured/calculated concentration, and U_{ref} is the reference velocity. Figure 4a shows the geometric setup and Figs. 4b–d show examples of model outputs for Cases 1 and 3 together with comparisons with measurements from controlled wind tunnel experiments. In detail, Fig. 4b shows the x -velocity component (along the wind direction) contour obtained from CFD simulations, where it

Urban Air Quality: Meteorological Processes. Table 5 Summary of the single building case. H , W , and L are the building height, width, and length, respectively. $H = 0.3$ m for case 4 and $H = 0.1$ m for all other cases

Case		Model details	Source details
1	Isolated rectangular building	Cube	The stack, 0.5 H high, is installed at an upwind distance of 2 H from the building
2		$W/H = 2$, $L/H = 1$	
3		$W/H = 1$, $L/H = 2$	
4		$W/H = 1/3$, $L/H = 1/3$	
5		$W/H = 6$, $L/H = 1$	



Urban Air Quality: Meteorological Processes. Figure 4

(a) Sketch of the single building with indication of the upstream source position and profiles where wind tunnel and CFD results are compared. (b) Contour of the x-velocity component in the middle of the building from CFD simulations. (c) Concentration profiles for Case 1. (d) Concentration profiles for Case 3

can be noted that the k- ϵ model was successful in predicting the typical vortex which develops in the wake zone behind the building. Figs. 4c and 4d also show the comparison between wind tunnel and CFD data for Case 1 and Case 3. In particular, it can be noted that in Case 1 the model under-predicts concentrations slightly. The hit rate q , a statistical quantitative score of the model performance [71], is equal to 75%. It is recalled that the Hit Rate validation test is performed using a fractional deviation $RD = 0.25$ and an absolute deviation $W = 0.06$ ($q > 66\%$ is typically requested for the comparison with wind tunnel data). Nowadays, statistical metrics are typically used to assess model performance during the model evaluation process. Usually k- ϵ models give a satisfactory performance for wind profiles according to the q tests. For this study, the Case 2 and Case 3 model predictions are in general agreement with observations from wind tunnel measurements, with hit rate test scores of $q = 82\%$ and $q = 77\%$, respectively. Finally, Case 4 and Case 5 model predictions are also in general agreement with observations from wind tunnel measurements with hit rate test scores of $q = 78\%$ and $q = 68\%$, respectively.

Isolated Street Canyons

Recent interest has risen in employing LES models ([72], [73]) to overcome the shortcomings of RANS in its inability to capture the unsteady and inherent fluctuations of the flow field within the street canyon on which the dispersion of pollutants depends. RANS, which is the most widely used approach in industry for the modelling of turbulent flows, assumes that non-convective transport in a turbulent flow is governed by stochastic three-dimensional turbulence possessing a broadband spectrum with no distinct frequencies and, therefore, models the entire range of the eddy length scales. This approach has obvious weaknesses and poses serious uncertainties in flows for which large-scale organized structures dominate, such as flows around buildings. In addition, RANS models assume gradient transport, which may not be the case for pollutant exchange at the roof level of a street canyon. LES, although computationally more expensive, has an advantage over RANS in that it explicitly resolves the majority of the energy carrying large-scale organized structures and the internally or externally

induced periodicity involved, whereas only the universal small-scale eddies are modelled.

As an example, an illustration is provided of the flow structure and pollutant dispersion within an urban street canyon of width to height ratio $W/H = 1$ using the standard k- ϵ model, the Reynolds Stress Model (RSM), and LES, coupled with the advection-diffusion method for species transport [74] (Fig. 5). In the case shown, the computational domain was built by using about one million cells with smallest dimensions equal to $0.077 H$. Numerical results, which include the statistical properties of pollutant dispersion, for example, the mean concentration distributions, three-dimensional spreads of the pollutant, etc., are then compared to wind-tunnel measurements from the online database [75]. Integrated into the model street, four tracer gas emitting line sources were used for simulating the release of traffic exhausts. Mean concentrations are normalized according to:

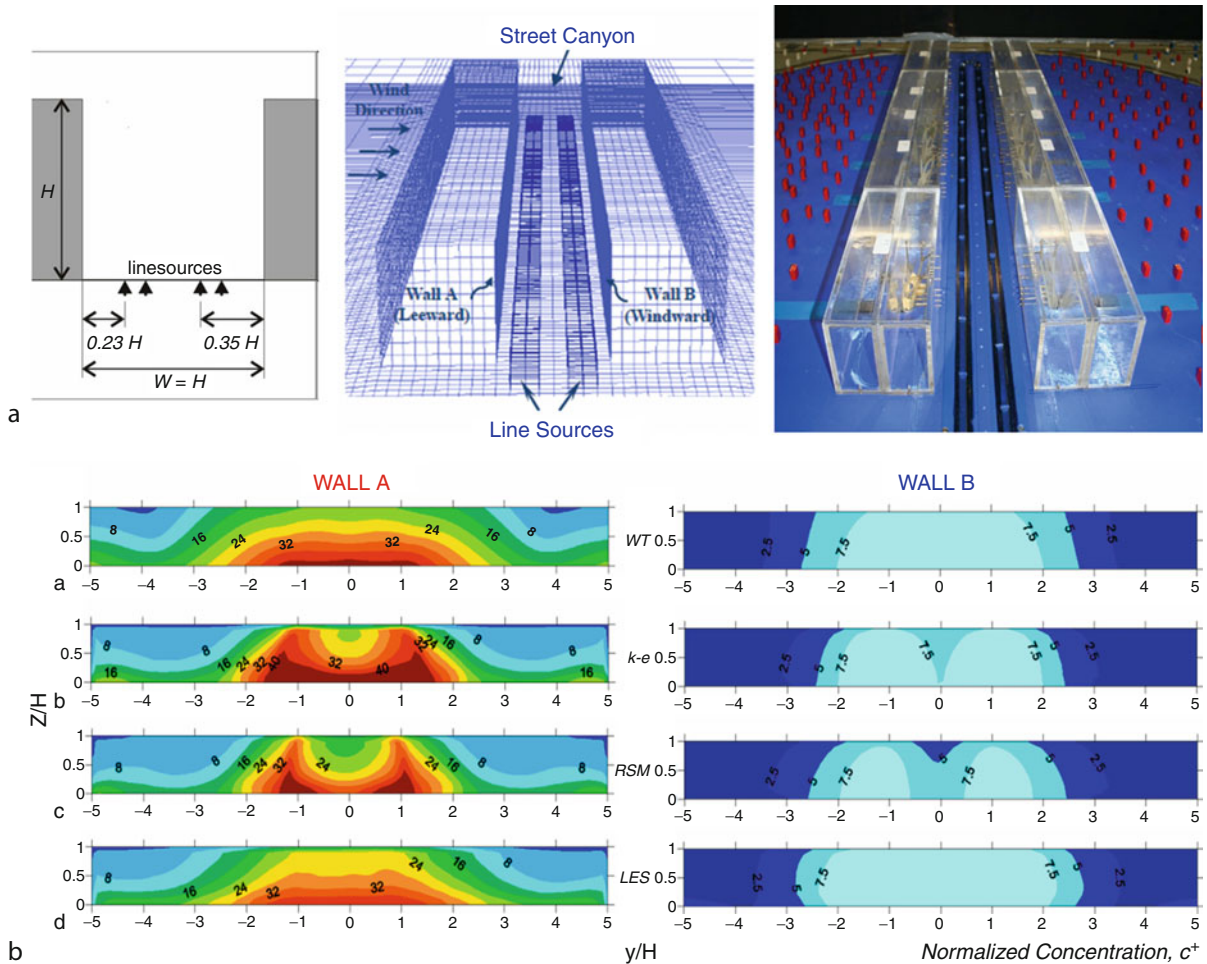
$$c^+ = \frac{CU_{\text{ref}}H}{Q/l},$$

where C is the measured/calculated concentration, H the building height, U_{ref} the flow velocity at height H in the undisturbed approaching flow, and Q/l the tracer gas source strength per unit length. Figure 5 shows normalized concentration contours at wall A (leeward) and wall B (windward) of the street canyons.

It is observed that amongst the two RANS models, RSM performed better than standard k- ϵ , however, LES proved better than RANS in predicting the concentration distribution because it was able to capture the unsteady and intermittent fluctuations of the flow field, and hence, resolve the transient mixing process within the street canyon.

Regular Building Arrays

Wind tunnel experiments have demonstrated that street ventilation is reduced in the presence of upstream buildings. This seems to be because of the upward displacement of the flow and the consequent perturbed momentum exchange between the street canyon and the outer region of the flow. Also, numerical results have indicated that the surrounding building configuration affects pollutant dispersion in a street canyon, therefore it should be taken into account in numerical



Urban Air Quality: Meteorological Processes. Figure 5

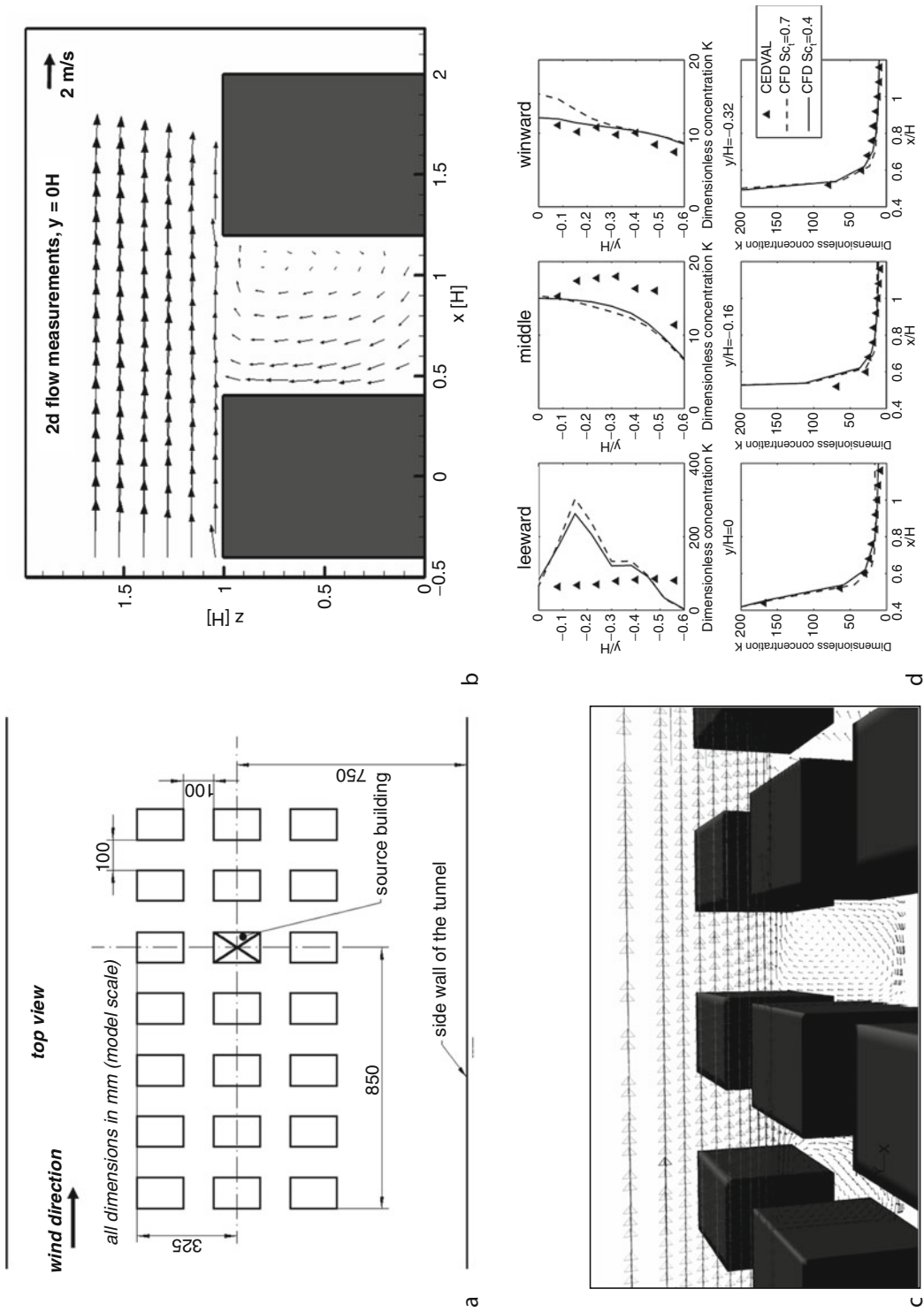
(a) Sketch of the street canyon with indication of the position of line source, grid, and wind tunnel model (from [74] CODASC, 2008). (b) Mean normalized concentrations at wall A (leeward) and wall B (windward) showing comparison between WT data (available from CODASC) and numerical results performed with $k-\epsilon$, RSM, and LES simulation

dispersion modelling (see, for example, [76]). Two examples of flow and pollutant dispersion distributions for regular building geometries are shown in Fig. 6, investigated by means of wind tunnel experiments and numerical simulations as previously described.

The experiment is set up in a boundary layer wind tunnel to simulate the case of dispersion of pollutants from naturally ventilated underground parking garages [77]. A finite array of idealized building blocks with 0.1 m by 0.15 m base dimensions and 0.125 m height (three buildings crosswind and seven buildings along wind direction) are considered, as shown in Fig. 6a. The aspect ratio of the street canyons resulting from the

building arrangement is $W/H = 0.8$. Four ground-level emission sources were mounted close to the building. RANS type simulations [78] were performed by employing the standard $k-\epsilon$ model and the advection-diffusion model. The computational grid was built by using 500,000 cells, with smallest dimensions equal to 0.005 m. Flow and dispersion were measured within the street canyon positioned downwind of the building equipped with the sources.

Wind tunnel experiments and CFD simulations show that a recirculation flow region forms within a street canyon exposed to a perpendicular flow. Both wind tunnel measurements and CFD predictions show



Urban Air Quality: Meteorological Processes. Figure 6

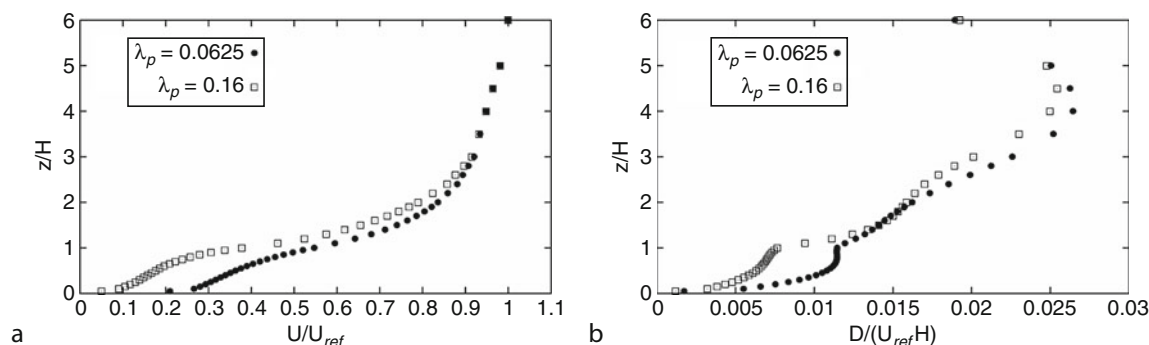
(a) Sketch of the regular building array with indication of the source building. (b) Velocity vectors obtained from wind tunnel experiments showing the vortex inside the canyon downwind of the source building. (c) Velocity vectors obtained from CFD simulations. (d) Concentration profiles at different positions within the same street canyon with a sensitivity to the Schmidt number value

that for aspect ratio $W/H = 0.8$ a single vortex forms inside the canyon. In particular, Fig. 6 shows velocity vectors on the vertical plane in the middle of the street canyon obtained from wind tunnel experiments (Fig. 6b) and CFD simulations (Fig. 6c). The figure shows that the shape and dimension of the vortex predicted by the CFD simulations is very similar to that observed by CEDVAL experiments [77] compiled at the University of Hamburg. The ability to capture the right dimension of the vortex is very important as the vortex drives the distribution of pollutants within the street canyon.

As outlined in [79] and [80], it should be noted that typical CFD codes predict dispersion spread that is smaller than that predicted by the well-validated integral model ADMS-Urban [81]. From a sensitivity test on the turbulent Schmidt number value (the standard value is 0.7), the value of 0.4 was found to be the most appropriate in order to artificially increase plume dispersion, even though such tuning should be avoided. Figure 6d (top) shows the dimensionless coordinate y/H versus the dimensionless concentration K near the floor ($z/H = 0.06$) within the street canyon downwind the source building and dimensionless concentration K versus dimensionless coordinate x/H on the same vertical plane (bottom). The horizontal velocity near the floor is negative because the vortex in the canyon is clockwise. Thus, pollutants are carried toward the leeward side and mixed in the street canyon. CFD results agree well with wind tunnel measurements.

It is known that FCM type of models still suffers from their high computational cost which prevents them from being largely employed in operational applications. This has pushed the development of FAM models yet incorporating new understanding of flow and dispersion processes within the urban canopy layer. Recently, di Sabatino et al. [82] have proposed a new modelling approach for the computation of the spatially-averaged flow field, where the average is defined at the neighborhood scale, i.e., from 0.2 km up to 10 km. The underlying idea is the description of the drag forces in terms of height-dependent morphological parameters based on detailed knowledge of building geometry such as the planar area index λ_p and λ_f (introduced in section “Characteristic Regions of the Flow and Drivers”) and the reduction of the complete three-dimensional flow field to a one-dimensional quantity, given by the mean wind direction, and the three-dimensional spatial dependence to a much more simple dependence from the height z . As a consequence, only mean profiles are needed as input to the model. There are examples in the literature [83] showing the possibility of extending this modelling approach to an Eulerian dispersion model for the computation of the concentration field. Figure 7 shows the spatially-averaged profiles of wind velocity and diffusivity coefficient resulting from full CFD simulations.

This reduced description of the wind field and the diffusivity coefficient was used as input into the developed three-dimensional Eulerian dispersion model for



Urban Air Quality: Meteorological Processes. Figure 7

Spatially-averaged profiles of wind velocity (a) and diffusivity coefficient (b). H is the building (cube) height and U_{ref} the average velocity at the top

the dispersion simulation of a pollutant released from a point source. Results were then compared with CFD predictions.

Figure 8 shows concentration contours for the case $\lambda_p = \lambda_f = 0.16$. The pollutant source was placed along the x direction at a distance from the inlet of approximately 1/3–1/4 of the domain length and equidistant from both boundaries in the y direction. The domain height was 6 H , where H is the building height. The figure refers to a source height $Z_s = 0.5 H$ and shows the comparison at three horizontal sections at $z = 0.05 H$, $z = H$, and $z = 2 H$.

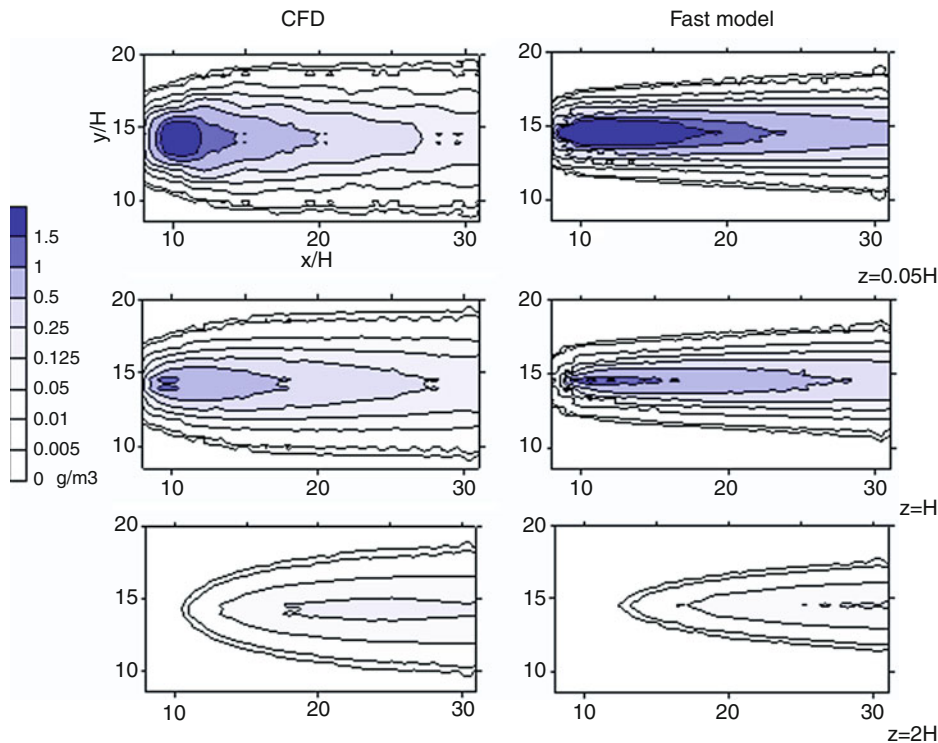
Inside the canopy, CFD concentrations tend to show some periodicity, due to the presence of cubic buildings. This specific behavior cannot be reproduced by the simplified Eulerian model. However, by using spatially-averaged one-dimensional profiles, the model is able to reproduce the order of magnitude of fully computational CFD concentration predictions, even if contour shapes may differ qualitatively. The quality of

the comparison is not homogeneous over the domain. As expected, the comparison is quite poor close to the source and improves downstream. Large differences are also found near the top boundary, but this is most probably an effect related to the boundary condition. Overall, the plume width predicted by the FAM type model is comparable to the CFD model. Similar results were obtained in the case of small packing density ($\lambda_p = \lambda_f = 0.0625$), not shown here.

Results highlight the potential of the approach based on the parameterization of one-dimensional wind and diffusivity profiles in terms of routinely available meteorological parameters and morphological parameters such as λ_p and λ_f , which synthesize the main geometric characteristics of the urban canopy.

Complex Urban Junction

The geometry of idealized building arrays is far less complicated than the morphology of real urban areas,

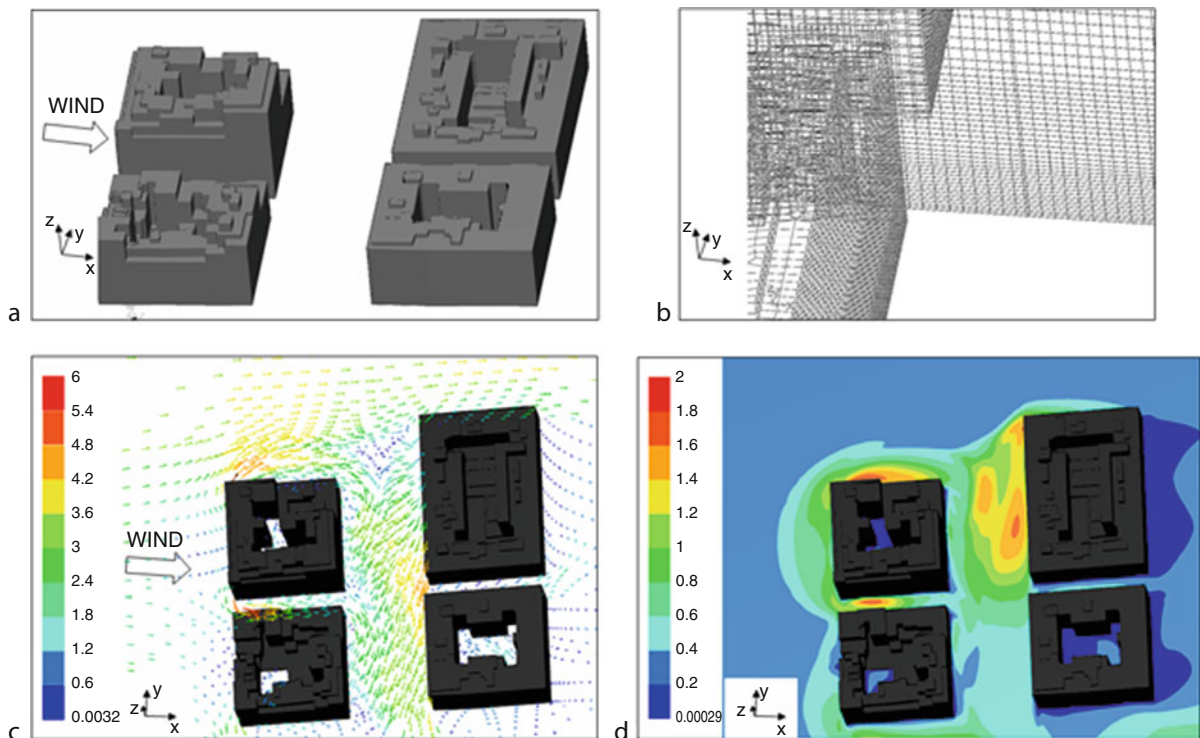


Urban Air Quality: Meteorological Processes. Figure 8

Intermediate canopy ($\lambda_p = \lambda_f = 0.16$), $Z_s = 0.5 H$: concentration (g/m^3) contours from the CFD model (left) and the fast model (right) at several horizontal planes

where urban neighborhoods are characterized by street intersections, asymmetric street canyons, and staggered arrays of buildings of irregular and/or highly variable shapes. These local features significantly alter the flow and dispersion patterns, as experimental and numerical studies have concluded (see for example [32], [84] and [85]). Reported here are some results of a modelling study performed with a CFD code by employing the RSM [86], and the advection-diffusion scheme in a neighborhood in the city of Bari in southern Italy. The real geometry comprises both step-up and step-down street canyons and consists of two asymmetric streets as shown in Fig. 9a. The wider street canyon is characterized by an average aspect ratio $W/H \sim 1.8$ and it is perpendicular to the wind direction. The other street canyon is characterized by an average aspect ratio $W/H \sim 0.5$ and is parallel to the wind direction. The tallest building is ~ 46 m. The two canyons intersect orthogonally forming four major blocks

of buildings and a junction. For the sake of computational convenience the real geometry was slightly simplified, yet maintaining the main geometrical details of the street canyons. A fine mesh close to the ground (up to a height of 4 m) was used (Fig. 9b), with cell dimensions $\delta x_{\min} = \delta y_{\min} = 1$ m, $\delta z_{\min} = 0.3$ m. The number of computational cells was approximately three and a half million. Vectors of velocity magnitude (m/s) and turbulent kinetic energy (m^2/s^2) at $z = 2$ m are shown in Fig. 9c, d, respectively. Due to the interaction of the wind with buildings, the resulting flow is channelled along the wider street canyon (which is that perpendicular to the wind direction), predominately blowing along the negative y direction. Obviously, the distribution of the turbulent kinetic energy is highly asymmetric, showing large values in the windward region. This example shows that flow pattern and turbulent kinetic energy are strongly affected by the geometry.



Urban Air Quality: Meteorological Processes. Figure 9

(a) Sketch of the urban junction. (b) Grid refinement close to the ground. (c) Flow pattern at $z = 2$ m developed within the urban junction. (d) Turbulent kinetic energy at $z = 2$ m

Future Directions

The entry has described the impacts of urban areas on urban meteorology at the mesoscale, neighborhood, and building street scale. While our general understanding of the processes is good and this enables us to construct approximate models, it is clear that at this time there is a paucity of detailed field data and the most detailed numerical models are still not able to resolve all the features of the flow in good time. With the availability of faster response and cheaper instrumentation, the continuing improvements in computer speeds, and the current strong interest in this topic, it is certain that more detailed data and more advanced models of urban meteorology will increasingly become available, resulting in huge data resources. Understanding of such data will continue to rely on the general concepts and categorizations described in this entry. Also of great relevance to future developments is both the impact of climate change on urban meteorology and the response of the urban environment to climate change. These issues are discussed in [87].

Notation and Abbreviations

b	Building breadth
C	Measured/calculated concentration
C_D	Drag coefficient
c⁺	Normalized mean concentration for a line source
d	Gap or separation distance between buildings
f	Coriolis frequency
F	Froude number
F_v	Surface water vapor flux
F_θ	F _θ
g	Gravitational acceleration
h	Boundary layer height
H	Building height
H_c	Canopy height
H_c'	Standard deviation of canopy height, H _c
H_M	Mountain height
K	Normalized mean concentration for a point source
k-ε	Kinetic energy and energy-dissipation model of turbulence

l	Vertical length scale of internal layer
l_O, l_N	Vertical length scales of internal layers over the urban area, neighborhood scale
L_A	Adjustment length for mean flow to adjust as it enters the porous canopy
L	Building length
L_f	Coriolis advection length
L_I	Inner city length scale
L_M, L_N, L_{BS}	Length scales of the (sub-)regions M, N, BS
L_O	Overall city length scale
L_{Ro}	Rossby length scale
L_{Se}	Effective source size
N	Buoyancy frequency
q	Hit rate test score
Q	Emission rate
RD	Fractional deviation
R_M	Ratio of the length of the sub-region L _M to the smallest scales resolved in that region
s	Distance to the nearest building
u_*	Friction velocity of the turbulent velocity profile of the atmosphere
U	Mean velocity, with subscripts denoting location/physical process
U_B	Typical wind speed associated with local buoyancy effects
U_C	Wind speed within the canopy
U_G	Geostrophic wind
U_H	Mean wind above the buildings (at height H)
U_{ref}	Reference velocity
U_c	Mean wind along the street canyon
V_S	Mean wind along street
w	Building length or width
W	Absolute deviation, Building width
x = (x, y, z)	Coordinates of a point
x_B⁽ⁱ⁾, y_B⁽ⁱ⁾	Coordinates of staggered building i
x_s, y_s, z_s	Coordinates of source
y_c(x)	Streamline through source at x _s , y _s , z _s
z₀(x, y)	Roughness length for wind profile
z_d	Displacement height for logarithmic wind profile
z_S(x, y)	Surface elevation of the ground
Z_s	Source height

Z_s	Height of top of shear layer above buildings
β	“Porosity” of an urban canopy [$\beta \sim bw/d^2$]
θ	Mean temperature
θ_s	Surface temperature
κ	Von Karman’s constant
λ_p	Planar area index
λ_f	Frontal area index
σ_u, σ_v	R.m.s velocity components (of the order of u_*)
ϕ	Angle between wind direction and normal direction to a street (Figs. 3 and 6), i.e., $\phi = 90^\circ$ if wind is along the street.

Subscripts

B	Buoyancy
BS	Building/Street scale
c	Canopy
C	Cloud concentration
f	Coriolis
G	Geostrophic
H	At top of buildings/canyon
M	Mesoscale
N	Neighborhood scale
O	Overall urban area
Ro	Rossby
s	Surface, street
S	Source
Se	Effective source
*	Turbulence-related level for log profile, or turbulent source

Abbreviations

BS	Building/street sub-region
CFD	Computational fluid dynamics
FAM	Fast approximate model
FCM	Fully computational model
LES	Large eddy simulation
M	Mesoscale region
N	Neighborhood sub-region
RANS	Reynolds averaged Navier–Stokes
RSM	Reynolds stress model
r.m.s	Root mean square
SVF	Sky view factor

Bibliography

- Hunt JCR, Carruthers DJ, Britter RE, Daish NC (2004) ADMLC, http://www.admlc.org.uk/documents/ADMLC20023_000.pdf
- Garratt JR (1993) The atmospheric boundary layer. Cambridge University Press, Cambridge, 334 p
- Hunt JCR, Simpson JE (1982) Atmospheric boundary layers over non-homogeneous terrain. In: Plate E (ed) Engineering meteorology. Elsevier Science, Amsterdam, New York, pp 269–318
- Oke TR (1978) Boundary layer climates. Methuen, London, 372 pp
- Mestayer PG, Anquetin S (1995) Climatology of cities. In: Gyr A, Rys FS (eds) Diffusion and transport of pollutants in atmospheric mesoscale flow fields. Kluwer, Dordrecht, pp 165–189
- Emmanuel R, Fernando HJS (2007) Effects of urban form and thermal properties in urban heat island mitigation in hot humid and hot arid climates: the cases of Colombo, Sri Lanka, and Phoenix, USA. *Clim Res* 34:241–251
- Högström U, Smedman A (1984) The wind regime in coastal areas with special reference to results obtained from the Swedish wind energy program. *Bound Lay Meteorol* 33:351–373
- Owinoh A, Orr A, Hunt JCR, Clark P (2003) Numerical modelling of boundary layer flow over a low hill with abrupt surface heat flux change (submitted)
- Belcher SE, Jerram N, Hunt JCR (2003) Adjustment of a turbulent boundary layer to a canopy of roughness elements. *J Fluid Mech* 488:369–398
- Grimmond CSB, Oke TR (1999) Aerodynamic properties of urban areas derived from analysis of surface form. *J Appl Meteor* 38:1262–1292
- Britter RE, Hanna SR (2003) Flow and dispersion in urban areas. *Ann Rev Fluid Mech* 35:469–496
- Kastner-Klein P, Rotach MW (2004) Mean flow and turbulence characteristics in an urban roughness sublayer. *Bound Lay Meteorol* 111:55–84
- Jackson PS (1981) On the displacement height in the logarithmic velocity profile. *J Fluid Mech* 111:15–25
- Davenport AG (1960) Rationale for determining design wind velocities. *J Struct Div Amer Soc Civ Eng* 86:39–68
- Davenport AG, Grimmond CSB, Oke TR, Weiriga J (2000) Estimating the roughness of cities and scattered country. In: Proceedings of 12th AMS conference on applied climatology, 2000. Asheville, NC, USA
- Cionco R, Ellefsen R (1998) High resolution urban morphology data for urban wind flow modeling. *Atm Env* 32:7–17
- Ratti CS, Di Sabatino S, Britter RE (2006) Urban texture analysis with image processing techniques: winds and dispersion. *Theor Appl Climatol* 84:77–90
- Macdonald RW, Griffiths RF, Hall DJ (1998) An improved method for estimation of surface roughness of obstacle arrays. *Atmos Environ* 32:1857–1864

19. Di Sabatino S, Leo LS, Cataldo R, Ratti C, Britter R (2010) On the construction of DEMs of a southern European city and a comparative morphological analysis with respect to northern European and North American cities. *J Appl Meteor Climatol* 49:1377–1396
20. Rotach MW (1995) Profiles of turbulence statistics in and above an urban street canyon. *Atmos Environ* 29:1473–1486
21. Counihan J, Hunt JCR, Jackson PS (1974) Wakes behind two-dimensional surface obstacles in turbulent boundary layers. *J Fluid Mech* 64:529–563
22. Mestayer PG, Vachon G, Rosant J-M (2002) The Nantes'99 data base for model validation of air quality in streets. In: *Proceedings of 8th International Conference on harmonisation within atmospheric dispersion modelling for regulatory purposes*, Sofia, Bulgaria, 14–17 October 2002, pp 297–300
23. Soulhac L (2000) Modélisation de la dispersion atmosphérique à l'intérieur de la canopée urbaine. Ph.D. thesis, École Centrale de Lyon
24. Todhunter PE (1990) Microclimatic variations attributable to urban-canyon asymmetry and orientation. *Phys Geog* 11(2):131–141
25. Kothari KM, Peterka JA, Meroney RN (1986) Perturbation analysis and measurements of building wakes in a stably-stratified turbulent boundary layer. *J Wind Eng Ind Aerodyn* 25:49–74
26. Lawson TV (1980) Wind effects on buildings, vol 1. Design, Applied Science
27. Britter RE, Hunt JCR (1979) Velocity measurements and order-of-magnitude estimates of the flow between two buildings in a simulated atmospheric boundary layer. *J Ind Aero* 4:165–182
28. Davidson MJ, Mylne KR, Jones CD, Phillips JC, Perkins RJ, Fung JCH, Hunt JCR (1995) Plume dispersion through large groups of obstacles – a field investigation. *Atmos Environ* 29: 3245–3256
29. Eames I, Hunt JCR, Belcher SE (2004) Lagrangian and Eulerian properties of steady flows through groups of obstacles. *J Fluid Mech* 515:371–389 [Summary version in *ERCOTAC Bulletin* No. 56]
30. Kastner-Klein P, Plate EJ (1999) Wind-tunnel study of concentration fields in street canyons. *Atmos Environ* 33:3973–3979
31. Schatzmann M, Leitl B (2002) Validation and application of obstacle-resolving urban dispersion models. *Atmos Environ* 36:4811–4821
32. Kastner-Klein P, Berkowicz R, Britter R (2004) The influence of street architecture on flow and dispersion in street canyons. *Meteorol Atmos Phys* 87:121–131
33. Macdonald RW, Ejim CE (2002) Flow and dispersion data from a hydraulic simulation of the MUST array. Technical Report 2002–2003, Department of mechanical engineering, University of Waterloo, Canada
34. Gryning S-E, Lyck E (1984) Atmospheric dispersion from elevated sources in an urban area: comparison between tracer experiments and model calculations. *J Clim Appl Meteorol* 23:651–660
35. Allwine KJ, Shinn JH, Streit GE, Clawson KL, Brown M (2002) Overview of Urban 2000: a multiscale field study of dispersion through an urban environment. *Bull Am Meteorol Soc* 83:521–536
36. Allwine KJ, Leach MJ, Stockham LW, Shinn JS, Hosker RP, Bowers JF, Pace JC (2004) Overview of joint urban 2003 – an atmospheric dispersion study in Oklahoma City. Symposium on Planning, Nowcasting and Forecasting in the Urban Zone, American Meteorological Society, Seattle, Washington, January 11–15
37. Rotach MW, Gryning S-E, Batchvarova E, Christen A, Vogt R (2004) Pollutant dispersion close to an urban surface – the BUBBLE tracer experiment. *Meteorol Atmos Phys* 87:39–56
38. Wood CR, Arnold SJ, Balogun AA, Barlow JF, Belcher SE, Britter RE, Cheng H, Dobre A, Lingard JJN, Martin D, Neophytou MK, Petersson FK, Robins AG, Shallcross DE, Smalley RJ, Tate JE, Tomlin AS, White IR (2009) Dispersion experiments in Central London: The 2007 DAPPLE project. *Bull Amer Meteor Soc* 90:955–969
39. Barlow JF, Dobre A, Smalley RJ, Arnold SJ, Tomlin AS, Belcher SE (2009) Referencing of street-level flows: results from the DAPPLE 2004 campaign in London, UK. *Atmos Environ* 43:5536–5544
40. Yee E, Biltoft CA (2004) Concentration fluctuation measurements in a plume dispersing through a regular array of obstacles. *Bound Lay Meteor* 111:363–415
41. Brown MJ, Williams MD (1998) An Urban Canopy Parameterization for Mesoscale Meteorological Models. Los Alamos National Laboratory, Report LA-UR-98-3831
42. WRF – <http://www.wrf-model.org/index.php> (2010) The weather research and forecasting model
43. Hodur R (1997) The Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). *Mon Wea Rev* 125:1414–1430
44. MM5-MM5 community model (2008) <http://www.mmm.ucar.edu/mm5/overview.html>
45. MESONH-Meso-NH, non-hydrostatic mesoscale atmospheric model (2011) <http://mesonh.aero.obs-mip.fr/mesonh/>
46. Cullen MJP (1993) The unified forecast/climate model. *Meteorol Mag* 122:81–94
47. Carlotti P (2002) Two point properties of atmospheric turbulence very close to the ground: comparison of a high resolution LES with theoretical models. *Bound Lay Meteor* 104:381–410
48. Hunt JCR, Orr A, Rottman JW, Capon R (2004) Coriolis effects in mesoscale flows with sharp changes in surface conditions. *Quart J Roy Meteor Soc* 130:2703–2731, Conference on shallow layer flows, Delft
49. Carruthers DJ, Hunt JCR, Weng WS (1988) A computational model of stratified turbulent airflow over hills – FLOWSTAR I. In: Zanetti P (ed) *Computer techniques in environmental studies*. Springer, Berlin, pp 481–492
50. Thykier-Nielsen S, Deme S, Mikkelsen T (1999) Description of the atmospheric dispersion module RIMPUFF. Report RODOS

- (WG2)-TN(98)-02. Risø National Laboratory. See also http://www.rodos.fzk.de/RodosHomePage/RodosHomePage/Documents/Public/Handbook/Volume3/4_2_6_RIMPUFF.pdf
51. Wippermann F (1984) Air flow over and in broad valleys: channeling and counter-current. *Beitr Phys Atmos* 57:92–105
 52. Blumen W (ed) (1990) Atmospheric processes in complex terrain. American Meteorological Society, Monograph vol. 23, Number 45, June 1990, 323 pp
 53. Ohashi Y, Kida H (2002) Effects of mountains and urban areas on daytime local-circulations in the Osaka and Kyoto regions. *J Meteorol Soc Jpn* 80:539–560
 54. Pope SB (2000) Turbulent flows. Cambridge University Press, Cambridge UK, 806 pp
 55. Hunt JCR, Olafsson H, Bougeault P (2001) Coriolis effects on orographic and mesoscale flows. *Quart J Roy Met Soc* 127:601–633
 56. Coceal O, Belcher SE (2004) A canopy model of mean winds through urban areas. *Quart J Roy Meteor Soc* 130:1349–1372
 57. Belcher SE, Xu DP, Hunt JCR (1990) The response of a turbulent boundary layer to arbitrarily distributed roughness change. *Quart J Roy Met Soc* 116:611–635
 58. Hunt JCR, Tampieri F, Weng W-S, Carruthers DJ (1991) Air flow and turbulence over complex terrain: a colloquium and a computational workshop. *J Fluid Mech* 227:667–688
 59. Murakami S (1992) Computational wind engineering one: Proceedings of the 1st International symposium on computational wind engineering (Cwe92). Elsevier Science, Tokyo, 21–23 August 1992, pp 924
 60. Rodi W (1997) Comparison of LES and RANS calculations of the flow around bluff bodies. *J Wind Eng Ind Aerodyn* 69–71:55–75
 61. Kim S-E, Boysan F (1999) Application of CFD to environmental flows. *J Wind Eng Ind Aerodyn* 81:145–158
 62. Boris JP, Grinstein FF, Oran ES, Kolbe RL (1992) New insights into large eddy simulation. *Fluid Dyn Res* 10:199–228
 63. Singh B, Hansen BS, Brown MJ, Pardyjak ER (2008) Evaluation of the QUIC-URB fast response urban wind model with a cubical building array and wide building street canyon. *Environ Fluid Mech* 8:281–312
 64. Moulinec C, Hunt JCR, Nieuwstadt F (2004) Disappearing wakes and scalar diffusion in tube bundles. *J Flow Turb Comb* 73:95–116 (See also ERCOFTAC Bulletin 56:5–9)
 65. Robins AG, Apsley DD (2000) Modelling of building effects in ADMS. ADMS technical specification, Paper P16/01 N/00. See http://www.cerc.co.uk/software/pubs/ADMS3-1TechSpec/P16_01.pdf
 66. Stoesser T, Mathey F, Froelich J, Rodi W (2003) LES of flow over multiple tubes. *ERCOFTAC Bulletin* 56:15–19
 67. Britter R, Schatzmann M (eds) (2007) COST Action 732, Background and justification document to support the model evaluation guidance and protocol. COST Office, Brussels
 68. Vardoulakis S, Fisher BEA, Pericleous K, Gonzalez-Flesca N (2003) Modelling air quality in street canyons: a review. *Atmos Environ* 37(2):155–182
 69. Wang X, McNamara KF (2006) Evaluation of CFD simulation using RANS turbulence models for building effects on pollutant dispersion. *Environ Fluid Mech* 6:181–202
 70. Launder BE, Spalding D (1974) The numerical computation of turbulent flows. *Comput Meth Appl Mech Eng* 3:269–289
 71. Schlünzen KH, Bächlin W, Brünger H, Eichhorn J, Grawe D, Schenk R, Winkler C (2004) An evaluation guideline for the prognostic microscale wind field models. Proceedings of the 9th International conference on harmonisation within atmospheric dispersion modelling for regulatory purposes, Garmisch-Partenkirchen (Germany)
 72. Cai XM, Barlow JF, Belcher SE (2008) Dispersion and transfer of passive scalars in and above street canyons—Large-eddy simulations. *Atmos Environ* 42:5885–5895
 73. Letzel MO, Krane M, Raasch S (2008) High resolution urban large-eddy simulation studies from street canyon to neighbourhood scale. *Atmos Environ* 42:8770–8784
 74. Salim SM, Buccolieri R, Chan A, Di Sabatino S (2011) Numerical simulation of atmospheric pollutant dispersion in an urban street canyon: comparison between RANS and LES. *J Wind Eng Ind Aerod* 99:103–113
 75. CODASC (2008) Concentration data of street canyons. Laboratory of Building- and Environmental Aerodynamics, IfH, Karlsruhe Institute of Technology. www.codasc.de
 76. Xie X, Huang Z, Wang J (2005) Impact of building configuration on air quality in street canyon. *Atmos Environ* 39:4519–4530
 77. CEDVAL (2001) Compilation of experimental data for validation of mesoscale dispersion models, <http://www.mi.uni-hamburg.de/Introduction.433.0.html>
 78. Di Sabatino S, Buccolieri R, Pulvirenti B, Britter R (2007) Application and validation of FLUENT flow and dispersion modelling within complex geometries. In: Borrego C, Renner E (eds) Air pollution modeling and its application XVIII, vol 6. Elsevier, Amsterdam, pp 3–11
 79. Riddle A, Carruthers D, Sharpe A, McHugh C, Stocker J (2003) Comparisons between fluent and ADMS for atmospheric dispersion modelling
 80. Di Sabatino S, Buccolieri R, Pulvirenti B, Britter R (2007) Simulations of pollutant dispersion within idealised urban-type geometries with CFD and integral models. *Atmos Environ* 41:8316–8329
 81. Carruthers DJ, Edmunds HA, Lester AE, McHugh CA, Singles RJ (2000) Use and validation of ADMS-urban in contrasting urban and industrial locations. *Int J Environ Pollut* 14:1–6
 82. Di Sabatino S, Solazzo E, Paradisi P, Britter RE (2008) A simple model for spatially-averaged wind profiles within and above an urban canopy. *Bound Lay Meteor* 127:131–151
 83. Di Sabatino S, Buccolieri R, Paradisi P, Palatella L, Corrado R A fast model for pollutant dispersion at the neighbourhood scale. *Int J Environ and Pollut*. Manuscript accepted for publication
 84. Wang J, Huang Z (2006) Numerical study on flow and dispersion in urban street canyons of asymmetrical configurations. *J Hydrodynamics* 18:146–150

85. Wood CR, Barlow JF, Belcher SE, Dobre A, Arnold SJ, Balogun AA, Lingard JJN, Smalley RJ, Tate JE, Tomlin AS, Britter RE, Cheng H, Martin D, Petersson FK, Shallcross DE, White IR, Neophytou MK, Robins AG (2009) Dispersion experiments in central London: the 2007 DAPPLE project. *Bull Am Meteorol Soc* 90:955–969
86. Launder BE, Reece GJ, Rodi W (1975) Progress in the development of a reynolds-stress turbulence closure. *J Fluid Mech* 68:537–566
87. Hunt JC, Bohnenstengel SI, Belcher SE, Timoshkina Y (2010) Implications of climate change for expanding cities worldwide. *J Urban Design and Planning*. Manuscript accepted for publication

Urban Air Quality: Sources and Concentrations

MICHAEL EVAN GOODSITE, OLE HERTEL

Department of Atmospheric Environment, National Environmental Research Institute, Aarhus University, Roskilde, Denmark

Article Outline

Glossary

Definition of the Subject and Its Importance

Introduction

Sources in Urban Airsheds

Impact of Physical Parameters: Geography, Topography, and Meteorology

Pollutant Dispersion in Urban Streets

Nitrogen Dioxide Pollution in Urban Areas

The Impact of Particle Pollution

Polycyclic Aromatic Hydrocarbons (PAHs) and Urban Air Quality

Trace Elements to Include Heavy Metals in Urban Areas

Conclusions

Future Directions

Bibliography

Glossary

Air pollutants Anything in the air around us man-made or natural that contaminates the natural steady state of and constituents in the air as required for life categorized as gaseous or particulate pollutants.

Air pollution The state of pollutants contaminating the air/atmosphere around us.

Air quality A subjective often locally or nationally determined state of the amount of pollutants in the air around us. The best air quality is when there are very low pollutant levels.

Air quality index A numerical index for air pollution, for example, in the United Kingdom it is from 1 to 10 and related to the air quality bands of “low,” “moderate,” “high,” or “very high” (http://www.airquality.co.uk/glossary.php?glossary_id=5, accessed February 13, 2010).

Ambient air Air outside of structures at any given time. The “outdoor air” as opposed to the “indoor air.”

Atmosphere The envelope of gases surrounding the Earth.

Coarse particles The coarse particle mass fraction is defined as the difference between measured PM_{10} and $PM_{2.5}$.

Diffuse sources of air pollutants Sources for which the release cannot easily be traced, for example, that are not geo coded like gases and particle emissions from smaller industry and domestic heating and ammonia from grassing cattle, etc.

Emissions The discharge of man-made or natural compounds into the environment.

Fine particles The $PM_{2.5}$ fraction of particles.

Particulates Fine and coarse liquid and solid particles of organic or inorganic substances present in the atmosphere/suspended (floating) in the air. Particulate matter (PM) consists of very small liquid and solid particles floating in the air.

Photochemical smog The resulting haze and reduced visibility from pollutants formed in reactions induced by sunlight and involving nitrogen oxides (NO_x) and hydrocarbons (HC). Photochemical smog including elevated ozone (O_3) levels occurs generally on warm sunny days that promote the photochemical reactions.

$PM_{2.5}$ The mass concentration of particles in the air less than 2.5 microns (μm) in diameter and thus capable of penetrating into the deepest parts of the lungs and therefore often those that can cause great health concern.

PM_{10} The mass concentration of particles $<10 \mu m$ in aerodynamic diameter. PM_{10} is generally used as an

indicator for suspended particulate matter (e.g., regulated in EU directives) and routinely measured at many locations throughout the world.

Point (stationary) sources of air pollutants Source that emit pollutants from one location, for example, power plants.

Smog A word combining “smoke” and “fog” first used by Dr. HA Des Voeux in 1905 describing the poor visibility and often sharp, unpleasant, and soot pertaining haze occurring during periods when air pollutants are trapped at ground level. There are two processes producing smog: photochemical and physical processes respectively referred to as photochemical smog and urban smog.

Ultrafine particles Particles in the size fraction below 0.1 μm in diameter. These particles are also often termed nanoparticles.

Urban An area with a large population, that is, town or city, generally with a high density and large amount of population.

Urban smog Smog where air pollutants particularly smoke and sulfur dioxide are trapped at ground level usually occurring on cold, wind-still winter days. The Great London Smog episode of 1952 is a historical example still seen in developing countries in today. In western countries, photochemical smog consisting of nitrogen oxides, ozone, and PAN is more common nowadays.

Winter smog Winter smog episodes including elevated levels of nitrogen oxides (NO_x) occur on cold calm days when air pollutants are trapped in urban areas by an inversion layer, that is, a layer of warmer air above the wind-still cold air.

Definition of the Subject and Its Importance

Urban air quality refers to how “clean” the ambient air is inside of cities with a density, population, and level of activity that generally are recognized as “urban.” Urban air quality generally differs from rural air quality since there are more concentrated sources and the ability for the pollutants in the air to be dispersed are limited by the physical constraints of the urban environment. Urban air quality varies significantly in urban cities in industrialized worlds, and nonindustrialized world generally because of the difference between control on pollution sources and use of open fires in the urban

areas. Nevertheless, illness, deaths, and damage to the environment are still attributed across the world to air pollution.

Worldwide, the United Nations Environmental Programme (UNEP) (http://www.unep.org/urban_environment/issues/urban_air.asp, accessed February 13, 2010) reports that an estimated more than one billion people annually are exposed to outdoor air pollution. Further, up to one million premature deaths are linked to urban air pollution in the ambient environment and about 1.5 million are linked to indoor air pollution. Given primarily these deaths and other health costs and loss of productivity, urban air pollution is estimated to cost approximately 2% of GDP in developed countries and 5% in developing countries.

UNEP further reports that rapid urbanization has resulted in increasing urban air pollution in major cities, especially in developing countries with over 90% of air pollution in major cities in developing countries attributed to vehicle emissions brought about by high numbers of older vehicles coupled with poor vehicle maintenance, inadequate infrastructure, and poor fuel quality with high amounts of sulfur in the fuel.

Most developed countries have put in measures to reduce vehicle emissions, in terms of fuel quality, for example, by regulating the amount of sulfur, benzene, and additives like lead in vehicle fuel and vehicle emission reduction technologies such as catalytic converters, these measures are yet to be adopted in most cities in developing countries and there are recent cases where vehicle manufactures save costs by not utilizing such technologies, while at the same time, lowering overall vehicle prices, thus enhancing the number of vehicles on the road with all of the subsequent effects on air quality.

Air quality as a term, is subjectively defined, generally through laws regulating the maximum concentrations of pollutants in the atmosphere via the maximum allowable emissions from sources. The quality of the ambient (outside) air is monitored, in some urban environments continuously with advanced equipment and generally reported in some type of index by local or national authorities; a so-called air quality index.

Despite relatively unchanged levels of pollutants, the air quality can be improved or degraded based on

the weather and topography of the urban environment. It is therefore a very complicated issue to manage and predict. Many municipalities therefore have air quality management (AQM) offices and/or plans.

There are specific mathematical models developed to predict air quality and support the air quality policies that naturally must be aligned with policies affecting the other environmental compartments.

In cities in the industrialized world, air quality has generally improved after controls and policies on pollution and emittants were emplaced during the past decades. The policies have been the result of international treaties, as well as national and local legislation. The United Nations Environmental Programme (UNEP) maintains a Register of International Treaties and Other Agreements in the Field of the Environment for the interested reader (http://www.unep.org/law/PDF/register_Int_treaties_contents.pdf). Specifically, the 1979 Convention on the Long-range Transboundary Air Pollution and its eight associated protocols, generally named after the cities in which they were signed such as Aarhus, Denmark are relevant for air quality as with the signing of the convention it is recognized as a transboundary issue. There is likewise an international agreement regarding transboundary haze pollution.

Air quality issues are generally on a long-term scale, however in urban areas, municipal authorities need to be prepared for acute events such as those that may be arising from an industrial accident such as what occurred at a pesticide plant in Bhopal, India in December, 1984, or from elevated levels of air pollution such as experienced in London during the “Great London Smog” of 1952. It has been estimated that just in this single event about 12,000 deaths could be directly attributed to the incident [1].

As point sources such as those responsible for the acute incidents above are addressed by policies, other man-made sources, generally the “mobile sources” (i.e., vehicular traffic) are becoming the new primary contributor to degraded air quality in urban areas.

Natural sources are also an important contributor depending on the locality with dust, sea spray, and other types of particulate matter (PM) such as pollen as well as volatile organic compounds (VOCs) emitted from plants and trees being those that contribute often to degraded air quality.

Historically in urban areas across the world, high levels of smoke and sulfur dioxide from domestic burning of coal in people’s homes and later combustion of fossil fuels in coal-fired power plants have been the major point source of air pollutants, and this has led to major episodes of air pollution with resultant degraded air quality and associated health effects, including deaths directly attributed to the poor air quality.

Air quality in urban environments, especially in developed countries is nearly always a function of vehicular traffic and the emissions arising from vehicles. Both petrol- (gasoline) and diesel-driven vehicles emit vast amounts of pollutants. These include carbon monoxide (CO), the oxides of nitrogen (NO_x), and volatile organic compounds (VOCs). Particulates (PM₁₀) and finer-sized particles (PM_{2.5}) are emitted not just from the combustion processes in the motor but also and more predominantly from physical processes, such as resuspension of road dust and wear of tires, brakes, and road material.

Introduction

Air pollution is estimated by WHO to cause about two million premature deaths (half of this is associated with outdoor and the other half with indoor air pollution) worldwide annually [2], and this is in addition to a variety of other adverse health effects like asthma and other airway diseases, cardiovascular disease, different types of cancer, as well as adverse pregnancy outcome. The most hazardous of the ambient pollutants are generally believed to be related to the particulate matter (PM), although the mechanisms behind these health effects are not yet fully understood. It is therefore still very uncertain what it is about the particles that cause these negative health effects. Despite the uncertainties related to the mechanism behind these effects, it has, for example, been estimated by WHO that reducing ambient air concentrations of PM₁₀ from 70 to 20 µg/m³ would lower the number of air quality related deaths by approximately 15% [2].

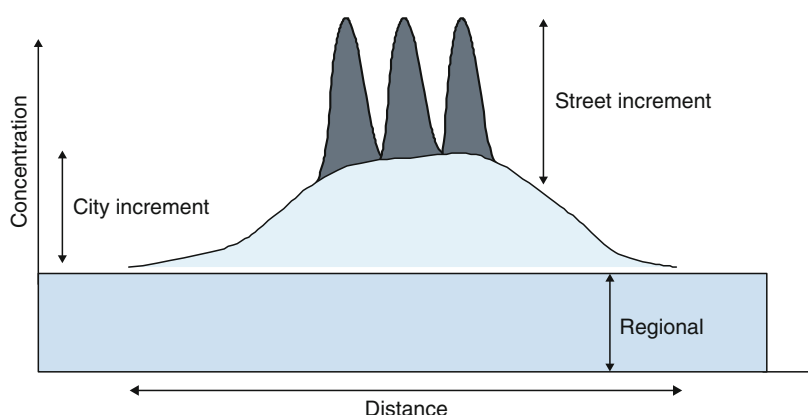
People are generally most at risk in the urban areas with respect to health effects related to air pollution. More than half of the world’s population reside in cities, and it is here that the highest air pollution exposure [3] and thus also most of the associated

negative health impact takes place. The projections for the next 50 years furthermore indicate that the world-wide urban population will be increasing by two thirds [3]. As a result of rapid urbanization, the urban air pollution has been increasing in many major cities, especially those found in developing countries (such as in Brazil, Russia, India, Indonesia, and China). The cost to society of the associated health effects is very significant, and has been estimated to be approximately 2% of the gross domestic product (GDP) in developed countries, whereas it is even 5% of the GDP in developing countries (www.unep.org/urban_environment/issues/-urban_air.asp). There may also be significant associated losses in productivity [5].

Origins of Air Quality: Sources and Formation of Urban Air Pollution

Urban ambient air pollution concentrations arise from the continuous competition between the emission processes that increase pollutant concentrations, and the dispersion, advection, and deposition processes that reduce and remove them. Pollutant differences between different urban areas are naturally reflecting differences in emission densities and pattern, but also in the dispersion and removal processes. The local meteorological conditions are governing the impact on pollution

levels of these dispersion and removal processes, and these meteorological conditions vary heavily with the physical location of the city. Air pollution concentrations in an urban environment are naturally the result of the contribution from local emissions, but also contributions arising from pollution from more remote sources may play a very important role (Fig. 1). The size of the city domain together with the density of pollutant emissions governs the local contribution to urban air pollution [6]. Naturally, the temporal pattern in urban air pollution levels is a function of the temporal pattern in the local releases. However, the extent of the urban area, and the spatial distribution in the local emission density, also play central roles for the local contribution to air pollution levels in the urban environments. The local contribution to pollution levels also varies heavily from one pollutant to another. In addition, the meteorological conditions greatly affect the actual pollution levels in a given situation as they govern the dispersion conditions and the pollutant transport in and out of the city area. In this context, the presence of the building obstacles plays a crucial role in causing generally high pollutant levels in the urban environment. This is especially true inside the street canyons with buildings on both sides of the street and where the canyon vortex flow governs the pollution distribution.



Urban Air Quality: Sources and Concentrations. Figure 1

City cross section. A schematic illustration of the air pollutant contribution from regional transport, the city area, and the street. The relative magnitude of the various contributions depends on the considered pollutant and the actual dispersion conditions (governed by the meteorology)

Air Quality and Health Effects

Air quality and health effects have been shown to be linked. Particulate matter (PM) is generally believed to be the most hazardous of ambient pollutants. More than half of the world's population reside in cities and the number is expected to grow [3]. Cities and especially their urban centers are where air quality is generally the poorest, and the highest air pollution exposure [4] and associated negative health impact therefore take place.

With projections for the next 50 years indicating that the worldwide urban population will increase by two thirds [3], urban air quality will continue to deteriorate as air pollution continues to increase in major cities, especially those in developing countries (such as in Brazil, Russia, India, Indonesia, and China).

Urban Air Pollution Concentrations in General

Urban air pollution arises from the competition between emission processes that increase pollutant concentrations, and dispersion, advection, and deposition processes that reduce and remove them. Differences in pollution levels reflect differences in emission densities and emission patterns. They also reflect processes of pollutant dispersion and removal. Pollution is dispersed by the flow of wind. It is removed generally via chemical conversion to other compounds and/or by being deposited either through wet or dry deposition processes.

The dispersion and removal processes are governed by the local meteorological conditions, which also vary heavily with the physical location of the city. Therefore, air pollution concentrations in an urban environment are naturally the result of local emissions as well as contributions from pollution transport from more remote sources (Fig. 1).

Also of importance to the local contribution to urban air pollution are demographic/geographic factors such as the size of the city domain and the density of pollutant emissions inside the domain [5].

The temporal pattern in urban air pollution levels varies as a function of the local releases. However, it is the variations in the meteorological parameters that govern the dispersion and also the pollutant transport into and out of the city from surrounding rural areas.

The temporal variations in emissions and meteorology are not the only contributors, the emission release height also plays an important role with respect to how quickly the air pollution is dispersed and deposited, with air quality generally improving with areas with very tall and technologically up-to-date smoke stacks. This is because air pollution emitted from a high release height will in many cases be transported out of the urban area before being dispersed down to ground level, although this is of course depending on the size of the urban domain. This is the principle that the "solution to pollution is dilution" when often, the pollution and subsequent air quality issues are just moved from one jurisdiction to another.

Urban industries, power plants, and other sources for which the releases come from tall chimneys therefore only rarely contribute to the local ground level air pollutant concentrations inside the urban area, and instead they are usually contributing primarily to the regional air pollution.

Pollutant emissions related to vehicular transport, local domestic heating, and smaller industries have low release heights [less than 10 m above ground level (a.g.l.)]. Consequently, the releases are not diluted as efficiently before reaching ground level as in the case for emissions from tall (more than 20 m a.g.l.) release heights. The contribution from "low" sources, both point and mobile, therefore often dominates the pollutant concentrations at ground levels inside the urban area.

A steady growth in vehicular transport and centralization of domestic heating have made road traffic the most important source of urban air pollution in many countries including most of the industrialized nations [7].

There are usually significant differences between developed and developing countries when comparing emissions from different industrial sectors. With respect to the local contribution from the various sectors, a comparison of two so-called megacities (Beijing and Paris) showed that aerosol particles and volatile organic compounds (VOCs) have a complex and multi-combustion source in Beijing, whereas a single traffic pollution source completely dominates the urban atmospheric environment in Paris [8]. Both municipalities recognize the above and are taking steps to improve air quality.

Urban Air Quality: Sources and Concentrations. Table 1 Ranges in annual average urban ambient air concentrations ($\mu\text{g m}^{-3}$) of PM_{10} , NO_2 , SO_2 , and 1 h average maximum concentrations of O_3 for different regions, based on a selection of urban data from WHO report [1] and previously presented by Hertel and Goodsite [66]

Region	Annual average concentrations			1 h max concentration
	PM_{10}	NO_2	SO_2	O_3
Africa	40–150	35–65	10–100	120–130
Asia	35–220	20–75	6–65	100–250
Australia/New Zealand	28–127	11–28	3–17	120–310
Canada/United States	20–60	35–70	9–35	150–380
Europe	20–70	18–57	8–36	150–350
Latin America	30–129	30–82	40–70	200–600

Source: Reproduced from Hertel and Goodsite [66]

Although ambient air quality is that which is often the subject of research and the type of pollution that is heard about (and also the subject of this entry), indoor air quality cannot be forgotten. Since many people spend most of their time indoors, indoor air quality is a major health concern and may be worsened by the location of the office or home in the urban environment.

The issue, as with case of ambient air quality, is generally worse in the developing countries. Emissions from household use of fossil fuels in the year 2000 were estimated to account for 1.6 million deaths, mainly among women and children in the poorest countries [9] that are being exposed to elevated pollutant levels in the indoor environment.

The actual ambient air pollutant load and therefore average air quality greatly varies from one city to another, but, generally, major urban areas throughout the world have poor air quality, and, among these, the cities in the developing countries face the greatest challenges.

WHO has compiled a survey on typical ranges in ambient air concentrations of four indicator pollutants, summarized in Table 1.

Urban Air Quality Concentrations and Indices

The greatest levels of pollutants like PM_{10} and SO_2 are presently found in urban air concentrations in Africa, Asia, and Latin America, whereas, the highest levels of secondary pollutants (i.e., produced via

reaction in the atmosphere with primary pollutants) like O_3 , PAN, and NO_2 are observed in Latin America and in some of the larger cities and urban areas in the developed countries.

The environmental and human health impacts are particularly severe in megacities, which are cities of about ten million or more inhabitants [10].

Urban air pollution has thus become one of the main environmental concerns in Asia, and especially in China where the pollution load in megacities like Beijing, Shanghai, Guangzhou, Shenzhen, and Hong Kong is substantial and air quality can be greatly affected (Photo 1).

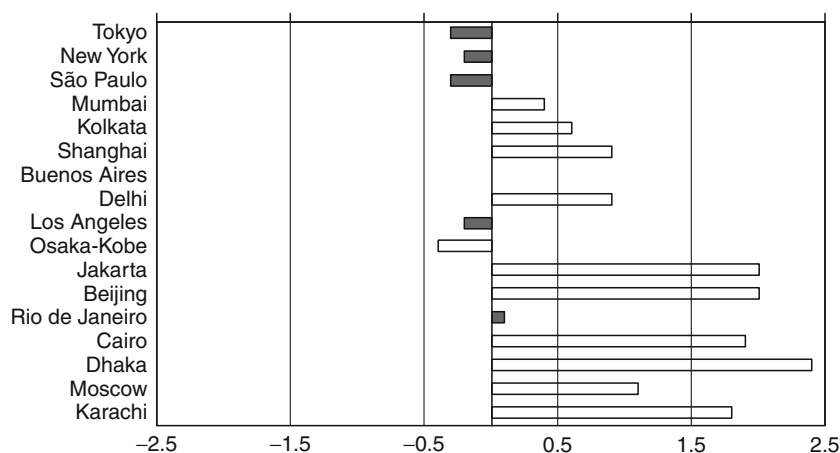
In these cities, between 10% and 30% of days exceed the so-called Grade-II national air quality standards [11] by a factor of three to five times that of the WHO AQG (air quality grade). These cities experienced a 10% growth in traffic each year over the last 5–6 years. Therefore even with enhanced emission controls, NO_2 and CO concentrations have remained almost constant over the same period of time.

Air quality indices (AQIs) are commonly used as tools in air quality management. A description of widely used indices and how they are expressed mathematically is given in Gurjar et al. [12]. AQIs may be designed to handle single or a multitude of pollutants, and they may also be used for comparing the loads in different municipalities or describing the current load in relation to average loads or air quality standards and target values.



Urban Air Quality: Sources and Concentrations. Photo 1

Photo from a Beijing Hotel, October, 2008 showing reduced visibility due to the poor air quality (Photo by M. Goodsite)



Urban Air Quality: Sources and Concentrations. Figure 2

Megacities pollution indices (MPI) based on measurements of the classical air pollutants and aggregated into an index for total pollution level (multipollutant). The plot is reproduced from Gurjar et al. [12], and 30% of the local NO_x contribution is related to aircraft, whereas the remaining 95–70% is from road traffic [15]

In an example of a multi component AQI (the authors applied the term MPI) a comparison over megacities throughout the world showed that the highest MPI values were found for Dhaka, Beijing, Cairo, and Karachi with values about double those of Delhi, Shanghai, and Moscow [12] (Fig. 2).

Sources in Urban Airsheds

Urban airsheds are the areas in and around the urban areas where air mass generally reaches a steady state given that they are facing stable, normal meteorological conditions.



Urban Air Quality: Sources and Concentrations. Photo 2

(a) and (b) Delhi in July 2010, immediately prior to rainfall at the Delhi airport and immediately after rainfall from the Delhi University Main Campus, showing the effectiveness of wet deposition in improving air quality (Photos by M. Goodsite)

There are many potential sources in urban areas that are generally located physically away from the urban center but still have to be considered as a part of the urban airshed. Airports are one of these.

Airports are usually located in the vicinity of larger cities and often mentioned as potential sources of high pollution loads in the urban areas (note: the haze in [Photo 2a](#) at the Delhi airport is probably not from the airport activities alone, but from the construction activities and traffic around the airport).

Several studies have thus been carried out to determine the potential impact of emissions from airport activities. These studies generally point at the main influence on air quality is coming from the road traffic going to and from the airport, whereas the impact of aircraft emissions has been found to be very limited. A study carried out in Frankfurt Airport, showed that aircraft-specific emissions generally could not be identified, whereas emissions from vehicle traffic on surrounding motorways had measurable impact on the air quality [13]. A study from Munich Airport had similar findings [14], as did a study inside Heathrow Airport [15].

Ship traffic has been estimated to be responsible for about 60,000 lung cancer and cardiopulmonary deaths annually [16], but this outcome is linked to the contribution from ship emissions to the

background particulate matter load and not directly related to contribution from ship emissions to the urban air quality.

Harbors may nevertheless be a local source contributing to urban pollution as they are often part of the urban airshed in coastal areas, but studies indicate that as with airports, that local road traffic often dominates the contribution from harbors. For example, a study in the harbor of Aberdeen thus showed a gradient of increasing NO_2 and soot concentrations from the harbor toward the city center [17], indicating that the contribution from the harbor had very limited impact on the local air quality in comparison with the emissions taking place in the urban environment.

Wood combustion in households is a concern in areas with many wood stoves. These areas have relatively high local emissions of PM in comparison with other anthropogenic pollution sources generally in the airshed. Investigations of wood combustion and air quality in developed countries like New Zealand [18], Sweden [19], the United States [20, 21], and Denmark [22] have documented that residential wood combustion likely significantly elevates the local PM concentrations in outdoor air. Emission inventories for Denmark, where wood-burning stoves are generally secondary sources of domestic heating, point at wood combustion as the largest anthropogenic source of

primary particle emissions in the country, being responsible for about 60% of the countrywide primary particle emissions.

Impact of Physical Parameters: Geography, Topography, and Meteorology

Impact of Geography

The location of the city has a great impact on the dispersion conditions and thus air quality as it affects the local meteorological conditions. Los Angeles for example, is situated in a valley with stagnant conditions during temperature inversions. The stagnant conditions lead to very low wind speeds, and this results in little air exchange between the valley and the surrounding areas. The hot and sunny climate and high emissions from traffic, industry, and other sources make the valley act like a large pollutant reaction chamber, and this is ultimately leading to high concentrations of photochemical products like ozone, nitrogen dioxide, and peroxy acetyl nitrate (PAN) with the resultant degradation of air quality.

Wind speed through a city plays an important role. For example, nitrogen oxide (NO_x) levels in the street Via Senato in Milan, Italy compared with the street Jagtvej in Copenhagen, Denmark show similar concentrations at the two sites despite the much higher traffic load in the street of Copenhagen [23]. This is directly the result of generally higher wind speeds in Copenhagen compared with Milan. High wind speeds and neutral conditions prevail in Copenhagen, whereas low wind speeds and stable or near stable conditions are frequent in Milan. Copenhagen has a cold coastal climate whereas Milan has a warm subtropical climate and the local wind conditions are furthermore affected by the location inside the Po Valley. Air quality planners must take their own geography into consideration.

Impact of Topography

Most cities have characteristic wind systems as a result of local topography. An example is the rising air over a warm mountainside during daytime often leading to local formation of clouds and the subsequent release of precipitation.

During the night, the system reverses and the cooling of the air in the mountain valley leads to stable conditions

that may lead to local air pollution problems. The impact of Katabatic winds is another example affecting cities along the Norwegian coast. Katabatic winds are formed when cold air masses move downslope (Katabatic is Greek for moving downhill) and meeting the colder snow- and glacier-covered areas, which then cool the air mass further as it continues flowing downhill.

Katabatic winds may lead, for example, to high levels of local dust with resulting air quality issues, though their occurrence is highly location specific. The warm and dry Foehn wind formed on the backside of a mountain chain, for example, on the north side of the European Alps is another case in point. As the wind flows over the mountain, the air mass is cooled and releases moisture. The air is subsequently warmed as the air mass flows downhill. This system may then form an inversion with the effect of reducing the dispersion of local air pollutants.

Impact of Meteorology

Ambient temperature in the urban atmosphere of larger cities is usually a couple of degrees Celsius higher than that found in the surrounding rural areas due to the so-called urban heat island effect [24]. The phenomenon is due to the much smaller reflection or albedo of the city as compared with its surrounding area and therefore absorbs more energy to release it later in the form of heat.

There is in addition a high consumption of energy inside the city, as a result of domestic heating and intense road traffic, which further contributes to the higher releases of heat than rural surroundings. Buildings act as heat reservoirs so the city has a less pronounced diurnal temperature variation compared with the rural area. Lastly, buildings and other urban constructions form a physical barrier for the wind flow. This shielding leads to less cooling of the surfaces inside the city.

In calm weather, an urban circulation cell, so-called heat island circulation, may be formed by warm air rising from the city. As it moves away from the city, the heated air may sink and then in another circulation, be returned to the city at a low altitude. A similar phenomenon is known in coastal regions, where a sea breeze may be formed as a result of the temperature difference between the sea and land surfaces.

A study in London showed that due to the heat island circulation over the city, the wind speed is never below about 1 m s^{-1} [25]. Therefore, the heat island effect is very important during low wind speed conditions in urban areas where it may dominate the dispersion and thereby be the limiting factor for the highest local air pollution concentrations and therefore local air quality. The urban heat island effect may thus be the limiting factor for the highest pollution concentrations and thereby air quality parameters in the urban environment.

Pollutant Dispersion in Urban Streets

Pollution “hot spots” are areas where large concentrations of pollution are emitted or concentrate. In urban areas, trafficked streets are air pollution hot spots (Fig. 1). The concentration inside the urban street are generally the result of two contributions, one from emissions from the local traffic in the street itself and one from background pollution entering the street canyon from above roof level [26]:

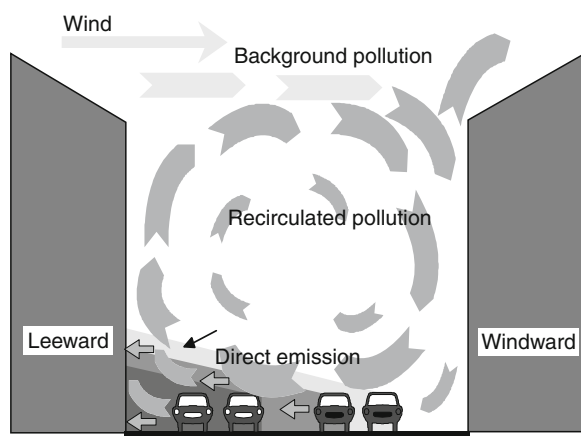
$$C = c_b + c_s$$

where c is the concentration in the street, c_b the urban background contribution, and c_s the contribution from traffic inside the street itself. The background contribution arises from two contributions: (1) the contribution from nearby sources in the urban area (typically this will be the traffic in surrounding streets) and (2) the regional (sources within a distance of a few hundred kilometers) and long-range transported (sources placed up to thousands of kilometers away) of air pollution.

The pollutant levels in the urban streets are strongly affected by the emissions from traffic taking place on the street itself. The concentration level and the distribution of air pollution inside the street canyon (the street “boxed in” by buildings on each side) are largely governed by the surrounding physical conditions. These physical conditions heavily affect the wind speed and the wind direction inside the street, with residence time for an air packet in the vicinity of an urban street usually on the order of seconds to a few minutes depending on the street topography. This means that only very fast chemical conversions of pollutants have time to take place [26].

The airflow generated inside the streets and around building obstacles may result in very different concentration levels at different places in the street.

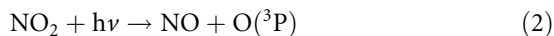
The street canyon vortex flow (Fig. 3) physically governs the pollutant distribution inside the street canyon. The street canyon is characterized by the presence of tall buildings on both sides of the street. Within the vortex, relatively clean air flows from rooftop height and down the windward face of the street, across the road at street level, in the reverse of the wind direction at roof top, thus transporting the pollutants from the road-level traffic to the leeward face of the canyon. The result is that pollution concentrations are up to ten times higher on the leeward side compared with the windward side of the street. This means that air quality depending on prevailing wind conditions may often be much poorer on one side of the street than another.



Urban Air Quality: Sources and Concentrations. Figure 3 Illustration of the flow and dispersion inside a street canyon. The wind above roof level blows perpendicular to the street in the above figure, thus creating a vortex inside the street canyon with the wind direction at street level opposite to the wind direction above roof level. The result of these flows is that there may be differences in air pollution concentrations of up to a factor of 10 between the two pavements inside the street canyon. Air quality on one side of the street might thus be poorer than on the opposite side of the street, if there are prevailing winds

Nitrogen Dioxide Pollution in Urban Areas

For example, the chemistry of nitrogen oxides [NO_x : the sum of nitrogen monoxide (NO) and nitrogen dioxide (NO_2)] in urban streets may be described by only two reactions: the reaction between ozone (O_3) and NO forming NO_2 , and the photo dissociation of NO_2 [27]:



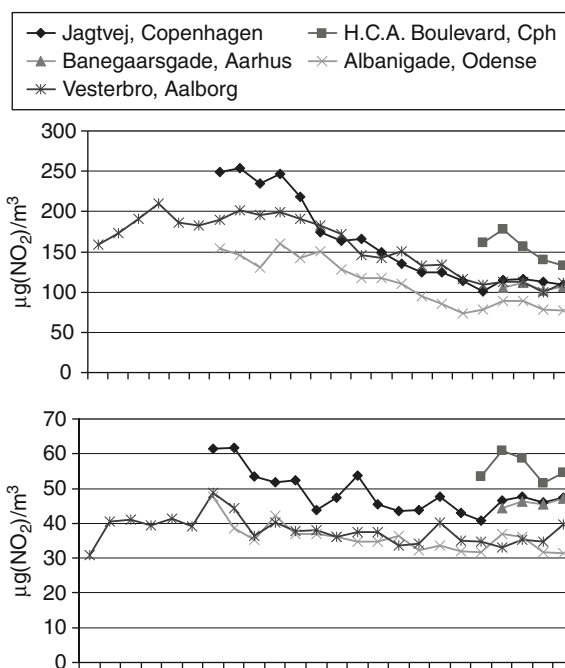
where $\text{O}(^3\text{P})$ is ground state atomic oxygen. Reaction (3) is very fast, and therefore for most practical applications it may be disregarded. The products of reaction (2) may thus be considered to be NO and O_3 .

NO_x is therefore mainly emitted as NO and to a lesser extent, NO_2 .

Long-term exposure to elevated NO_2 levels is a health hazard, and may decrease lung function and increase the risk of respiratory symptoms such as acute bronchitis, cough, and phlegm, particularly in children [28], whereas NO at current ambient air concentrations is considered to be harmless.

The fraction of NO_x directly emitted as NO_2 used to be only about 5–10% in countries with a small fraction of diesel engines. However, due to the use of catalytic converters and an increasing number of diesel engines with high fraction of NO_2 in the exhaust, this value may in some regions be now up to as much as 40% [29]. This extremely simple chemical mechanism of two reactions and a direct emission describes very well the concentrations of NO_2 inside urban streets [27], for many cities it may also be applied for describing NO_2 concentrations in urban background air (see [6] documenting this the case for many Northern European cities).

The use of catalytic converters in automobiles has in recent years led to a major reduction in NO_x concentrations in the urban streets of many industrialized countries. For several reasons, NO_2 levels have not followed the same trends (Fig. 4). Part of the explanation is the chemical conversion of NO to NO_2 in the reaction with O_3 , but the other explanation is, as mentioned above, the increased fraction of NO_2 in the NO_x



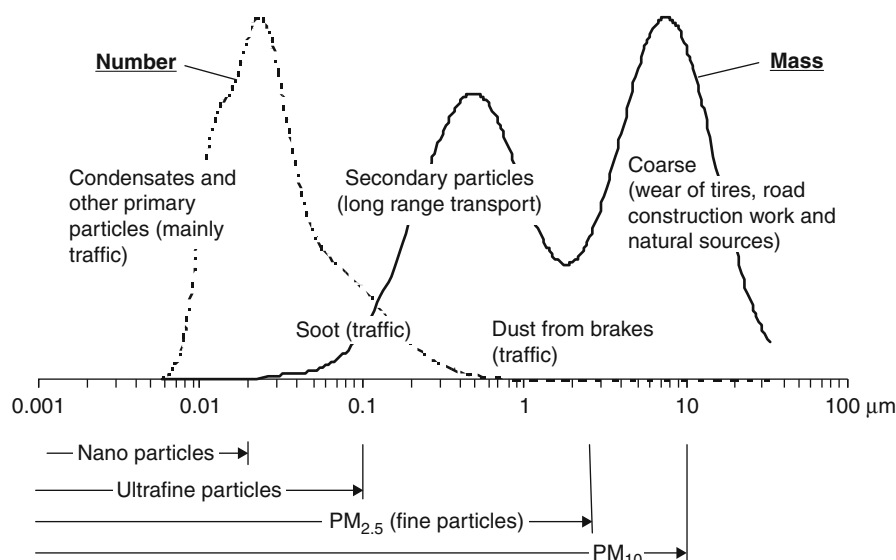
Urban Air Quality: Sources and Concentrations. Figure 4

The measured trend in annual mean concentrations of NO_x and NO_2 (both shown in $\mu\text{g}(\text{NO}_2)/\text{m}^3$) at the street stations in the largest Danish cities: Copenhagen, Aarhus, Odense, and Aalborg. The upper plot shows NO_x and lower plot NO_2 . The plots include measurements from the time period 1982–2005, and illustrate a decrease in NO concentrations resulting from increasing number of vehicles with catalytic converters in Danish car fleet, but also illustrates that this decrease is not reflected in the NO_2 concentrations that have remained more or less constant during this time period [64]

emission from vehicles with catalytic converters. Despite the overall reduction in NO_x emissions, this is contributing to elevated NO_2 concentrations.

The Impact of Particle Pollution

Ambient urban air contains a complex mixture of particles from natural and anthropogenic sources of varying sizes and chemical composition [30–34]. The particle size is crucial for the atmospheric fate [35] and the human health impact as both the particles' atmospheric behavior and their deposition in the human respiratory system is governed in part by its size [36].



Urban Air Quality: Sources and Concentrations. Figure 5

Typical size distribution of particles in urban air: both as mass and number concentration. The horizontal axis is the particle diameter in micrometers. The black line is mass distribution, dominated by the coarse and secondary particles. The dashed line is the number distribution, dominated by ultrafine particles. Note that one particle with a diameter of 10 μm has the same weight as one billion particles with a diameter of 0.01 μm [65]

Particles in ambient air generally appear in two rather distinct size classes (referred to as “modes”) usually termed fine particles (diameter: 0.1–2.5 μm) and coarse particles (diameter: >2.5 μm). In addition, it is common to talk about the ultrafine particles (diameter: 0.01–0.1 μm or 10–100 nm) that usually dominate the number concentrations (Fig. 5).

Particle Mass Concentrations

Mass concentration of particles <10 μm and <2.5 μm in aerodynamic diameter (PM_{10} and $\text{PM}_{2.5}$) are generally used as an indicator for suspended particulate matter (e.g., regulated in EU directives) and routinely measured at many locations throughout the world. Particle mass is dominated by particles >0.1 μm in diameter. Particles in traffic exhaust are mainly in the ultrafine fraction and include elemental carbon (EC) as well as organic carbon (OC) [37]. The ultrafine mode particles from traffic sources contribute considerably to the PM number concentrations but only little to the PM mass.

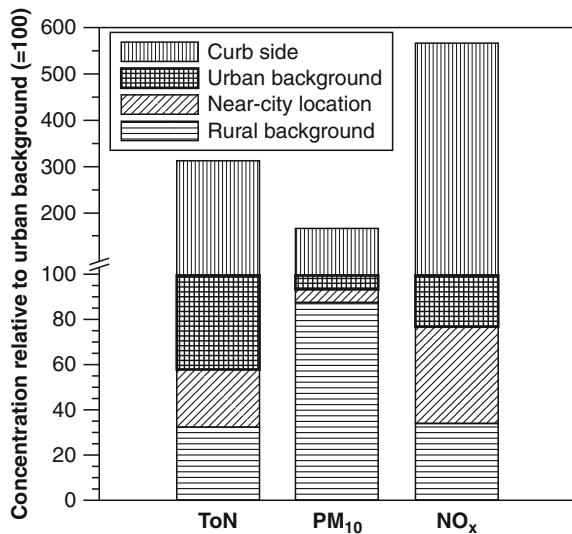
While some health studies have demonstrated a relationship between both acute and long-term health

effects and the ultrafine particle mode (particle diameters <0.1 μm) [38], other studies are less conclusive concerning the health effects of the ultrafine mode particles. Even though ultrafine particles are a minor contribution to total particle mass concentration, they represent most of the particles in terms of the number concentration. Most studies to this point of the health effects air particles have been related to particle mass. Neither of the mechanisms are yet fully understood and the need for more studies has been called for [39].

With respect to streets with a high level of vehicular traffic, a significant fraction of the particle pollution originates from this traffic [40]. The direct emissions from car exhaust contain particles formed inside the engine as well as in the air just after leaving the exhaust pipe. The latter depend on the sulfur content in the fuel.

Studies have shown that reducing the sulfur content in diesel significantly reduces the particle number concentrations in urban streets and this parameter is therefore important for air quality planning and management.

Directly emitted particles are mainly found in the ultrafine particle fraction of the total particle



Urban Air Quality: Sources and Concentrations. Figure 6 Comparison of average concentrations of total particle number (ToN), particle mass (PM₁₀), and NO_x at rural, near-city, urban, and curb side measurement stations relative to urban background levels in the area of Copenhagen, Denmark. The concentration bars are stacked so only the additional contributions are marked with the pattern shown in the legend. Note that the scale of the vertical axis changes at 100 (Adapted from Ketzel et al. [40])

concentration. Traffic, however, also contributes to the mechanically formed particles in the fine and especially the coarse fraction, typically produced from wear of tires and brakes, wear from road surface material as well as resuspended dust.

For the particle mass (PM₁₀), long-range transport is usually the dominating source for regional background levels. Studies have shown that less than 10% of the urban PM₁₀ originate from local urban sources [40]. For NO_x as well as particle numbers a larger difference between the urban, rural, and curb side levels have been observed, indicative of a large contribution from local traffic sources (Fig. 6).

Particle Number Concentrations

A high correlation (one study [40] found $R > 0.9$) between concentrations of NO_x and total particle number indicating that both components of air pollution originate from the same (traffic) source.

The correlation is seen as the components are emitted in a similar ratio (particle number : NO_x) from different traffic categories, in other words, high NO_x emitters such as diesel vehicles (especially heavy-duty vehicles) are also the high particle emitters (when these are expressed in particle number concentrations) [41].

The Danish Operational Street Pollution Model (OSPM) [42] is an example of a numerical model that through its calculations, have been shown to reproduce well the observed particle number, when treating particles as inert tracers (i.e., by disregarding transformation and loss processes) [43]. The particle emission factors depend on ambient temperature with higher emissions at lower temperatures, which is accounted for in the simulations reported in [44] and other refined numerical models.

Long-range transport however, is a significant contributor to fine fraction particles and leads to the main part of particulate sulfate and ammonium and a large part of particulate nitrate. These secondary particles are formed from anthropogenic sulfur dioxide (SO₂), ammonia (NH₃), and NO_x emissions [45] and often constitute more than 30% of the PM₁₀. A separate fraction of particulate nitrate appears in the coarse mode, which also contains contributions from sea spray and resuspended dust [46, 47] with relatively large mass and thus quickly deposits by gravitational settling, such as road dust. Due to the quick deposition, coarse particles, therefore, generally have a relatively short lifetime in the atmosphere when compared with fine particles.

Particles of low-temperature combustion processes from wood-burning stoves and other poorly controlled burning processes may be soot particles that often have high contents of polycyclic aromatic hydrocarbons (PAHs) and other carcinogens or just generally degrade the air quality in the urban environment, despite the “charm” or in developing countries, necessity, for burning with these methods.

Air Quality Measurement Locations

When comparing observed levels of particulate matter or air quality in different cities, aside from assuring that the indices are the same, the location of the monitoring stations must be taken into account, as they may be very different. An index that reports the “average”

might consist of a station in the urban center and one on the outskirts. There is a real risk of comparing sites in the vicinity of large pollution sources with sites at some distance from local sources. The location of measurement stations must be very carefully considered, and it is not generally good enough to just replicate what another town is doing. For all of the physical and anthropogenic factors previously discussed, locating stations is critical and must be done in conjunctions with city planners. Recently in the Danish press, scientists speculated if air quality levels improved as a result of newly laid road tar thus keeping down the dust at a roadway at the measurement station.

In addition to the ambient measurements, modern air quality management is starting to consider locating measurement stations where many people spend most of their time: indoors. Exposure to air pollution in the home is an important fraction of a person's overall exposure. A Danish study in an uninhabited apartment in central Copenhagen revealed that particle pollution inside the apartment was to some extent linked to the activity level of the neighboring apartments [48] indicating that sources in the neighboring environment must also be considered in the analysis of overall exposure. Measuring air quality is therefore not a simple matter of purchasing machines. They need to be carefully located and the measurements are best coupled with other data such as meteorology and with numerical models that help provide additional value with the measured numbers.

Natural Sources

A 2006 WHO study (Fig. 7) indicates that PM_{10} concentrations in Asia and Latin America are higher than are observed in Europe and North America [1]. The highest particle levels are observed in Asia and are attributed to forest fires, poor fuel quality, and aeolian (windblown) dust. Wind erosion originating especially in the deserts of Mongolia and China contributes to the general level of PM in the region.

Polycyclic Aromatic Hydrocarbons (PAHs) and Urban Air Quality

Sources and Emissions of PAHs

Polycyclic aromatic hydrocarbons are a group of chemicals that are formed during incomplete burning

of coal, oil, gas, wood, garbage, or other organic substances, such as tobacco and charbroiled meat. There are more than 100 different PAHs. PAHs generally occur as complex mixtures normally as part of combustion products such as soot and not as single compounds.

PAHs can also be found in materials such as crude oil, coal, coal tar pitch, creosote, and roofing tar. There are numerous other sources [49, 50].

In the air, they are either attached to dust particles or as solids in airborne soil or sediment. This is of great concern due to the mutagenic [51] and carcinogenic [52] properties of PAHs; the United States Agency for Toxic Substances (ATSDR, 1995) has listed 17 PAHs as of priority concern with respect to their toxicological profile.

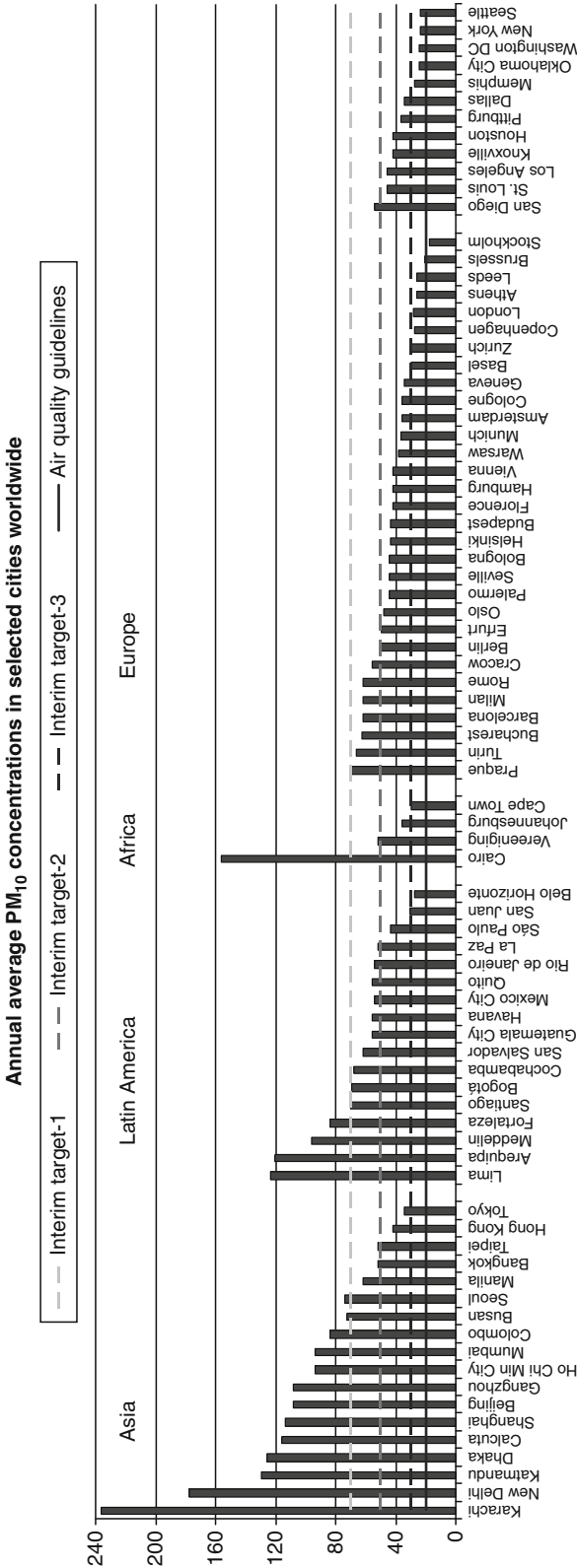
PAHs that have been studied in urban and other areas appear to represent only a fraction of a percent of the ambient particle mass [53]. However, there are compelling arguments that the previously commonly employed measurement techniques produced artifacts. Despite the toxicological profile of PAHs, no country has mandatory guidelines with respect to their ambient air quality standards [54].

Sampling Artifacts in Measuring PAHs

In order to reduce the risk of sampling artifacts [52], in the spring of 2003 the Mexico City Metropolitan Area (MCMA) air quality campaign employed three independent methods to measure particle-bound PAHs [57, 58]. Peak concentrations of PAHs on the order of 120 ng m^{-3} during the morning rush hour were observed in excellent agreement that 20% of the vehicles account for 50% of the PAH emissions [59]. Like other atmospheric gaseous and particulate constituents, PAHs are removed from the atmosphere by wet or dry deposition and may also be converted or degraded in heterogeneous processes. Despite the systematic approach, there may have been significant concentrations of very toxic and very reactive PAHs in the MCMA atmosphere that were missed due to filter reaction artifacts [58] thus attesting to how complicated measuring PAHs are as a part of air monitoring programmes and the need for further development.

Long-Range Transport of PAH

Together with the local emission, the long-range transport of PAHs is of concern. PAHs are designated as one



Urban Air Quality: Sources and Concentrations. Figure 7

PM₁₀ concentrations (μg m⁻³) in selected large cities throughout the world. The plot has been derived on basis of WHO (2006) [1]. The WHO Air Quality Guideline (AQG) (shown in the figure) is defined at the lowest level at which total, cardiopulmonary and lung cancer mortality has been shown to increase with more than 95% confidence in response to long-term exposure. Shown also are the WHO interim target values 1–3

of the persistent toxic substances in central and north-east Asia under the Stockholm Convention (UNEP, 2002). It has been demonstrated that transformation processes can lead to PAHs that are more toxic than their precursors [57].

The source apportionment of particulate PAHs at Seoul, Korea, during a measurement campaign between August 2002 and December 2003 were analyzed [60] by applying the US EPA (2004) chemical mass balance (CMB) model. Seoul is in the atmospheric footprint of major coal-burning industries and power plants of China and Japan, as well as contributing itself to the Northeast Asia particle footprint. In Seoul, similar to MCMA, gasoline and diesel vehicles accounted for 31% of the measured PAHs and thus were the major sources. Daily and seasonal variations were noted and attributed to differences in biomass burning and coal (heating) with a 19% difference in the total concentrations observed between fall (63%) and winter (82%). The sources showed an inverse profile when moving away from the city. The study authors used source analysis to attribute this to the long-range transport of atmospheric pollutant (LRTAP) of PAHs and their precursors from sources in China or North Korea thus demonstrating that for certain urban environments, in footprints of major industrial areas, PAHs from local and long-range transport must be taken into greater consideration as air quality management plans are updated.

Trace Elements to Include Heavy Metals in Urban Areas

Trace elements including heavy metals (HMs) are ubiquitous in the atmosphere of urban areas and many are classified by authorities as hazardous air pollutants (HAPs).

These represent serious environmental and health risks and are especially of concern in urban and industrial climates, with megacities in developing nations being the obvious sites of concern for populations at risk for trace element exposure.

Air quality can be affected and trends in urban areas have been noted. One study has provided trends and levels of trace metals in three Danish cities as well as at background sites [61]. The levels reported are less than

literature values for megacities, as expected, but in many cases the trends are similar, though the numbers of studies are very limited.

Heavy Metals in the Urban Atmosphere

Heavy metals are still a problem for ambient and indoor air quality, especially in general in urban and industrial areas despite that many nations have tightened and continue to restrict emissions of such trace elements as mercury. Even with history of progressive regulation of trace metals, they still represent a real and pressing health and environmental hazard in the United States (US EPA, 2006) and most certainly in other western nations and other countries as well. Trace metals have been measured in nearly all aerosol size fractions so it is of paramount importance to characterize the particle size distribution and relate these to the potential adverse health effects in the urban population [62].

Humans are exposed to metals via ambient air inhalation, and consuming contaminated food or water, as well as, in the case of lead and children, chewing on lead-painted toys, and exposure from walls or furniture. Metals were of great concern in western cities prior to lead being banned as an anti-knock agent in gasoline in the mid-1970s, as well as other regulations regarding the use of lead in, for example, house paints.

The United States Environmental Protection Agency (US EPA) lists many trace metals as among the worst urban air toxics (www.epa.gov/ttn/atw/nata/34poll.html, November 28, 2008) and several metals aside from mercury and lead are classified by the United States Clean Air Act as HAPs: chromium, manganese, nickel, and cadmium. There is no reason that other countries should not also endorse this classification in their air quality plans.

Trace Elements in the Urban Atmosphere

Trace elements (and metals) and the observed high levels of them in the urban environment have led to extensive monitoring with consequential regulations and international agreements that have greatly reduced the concentrations.

The sources of emission into the urban atmosphere of commonly measured trace elements: Be, Co, Hg,

Mo, Ni, Sb, Se, Sn, and V, with smaller concentrations of As, Cr, Cu, Mn, and Zn, are primarily human activities, especially the combustion of fossil fuels and biomass, industrial processes, and waste incineration [62].

As with PAHs, there are seasonal variations of trace metals in PM₁₀ and PM_{2.5} observed in campaigns conducted seasonally. Higher winter values of nearly all trace elements are observed suggesting a significant source of particles from domestic heating in most temperate urban areas, and/or less efficient dispersion of emissions in winter.

Many trace elements are primarily observed from natural sources and have local, regional, and global effects. Sources include sea spray near coastlines, dust from aeolian processes and weathering (especially noted in Asia), and volcanoes.

The main source of trace elements in most urban environments are nevertheless, anthropogenic activities, and national regulations, especially with regard to limits in gasoline, and emissions from coal-fired power plants, as well as other industrial control are in most cases effective in mitigating their emissions. Thus, developing nations generally have higher levels in general than industrialized nations.

Conclusions

Urban air quality is subjectively determined based on levels mandated generally by law. It is a function of a complex and dynamic mixture of gaseous and particulate pollutants emitted from natural and human environments and their interaction with natural and built environment. The concentration of pollutants both produced in photochemical reactions and suspended in the air of the urban environment and therefore air quality is weather dependent and thus exhibits daily and seasonal variation due to both anthropogenic and biologic activity levels and weather conditions.

The actual pollution load in a given urban area is the result of both local conditions and emissions and transport from nearby and remote sources. The location of the city is very important for the local dispersion conditions, which are governed by meteorology but are also heavily affected by topographical conditions (e.g., effects of coastline, mountains, and valleys).

Studies of the impact of transportation hubs such as local harbors and airports on urban air quality have generally pointed at a limited influence, with the impact on the air quality mainly related to the road traffic that these facilities are generating.

The highest urban air pollutant concentrations of the pollutants PM₁₀ and SO₂ are generally found in developing urban environments such as those in Africa, Asia, and Latin America.

The highest levels of photochemical pollutants such as O₃ and NO₂ are observed in Latin America and certain developed countries – topographical conditions and weather conditions are very important factors in this respect.

The negative health effects of urban air pollution are well documented and are found to be particularly severe in the megacities, where quality of life is lessened due to air pollution. The interaction of poor air quality with other lifestyle decisions such as smoking generally results in a synergistic effect. The negative health effects of poor ambient air quality are likely enhanced by exposure to tobacco smoke, indoor particulate and gaseous sources from cooking, candles, stoves, etc., but also by exposures to the non-airborne agents in textiles, food, etc. The possibility of reduced productivity from toxics and their associated costs should be investigated and valuation models should be further developed [5].

Even though the health effects of poor urban air quality have been documented in numerous studies, there are still many unknowns and through their investigation, there is a possibility for improving health, environment, and the climate. The reader is referred to Hertel and Goodsite, 2009 [66] – from which this entry was based; as well as the supplemental literature list where recent reviews across various aspects of the field of urban air quality are included.

Future Directions

The European programme Integaire (see www.integaire.org; Integaire was established in March 2002 under the key action “City of Tomorrow and Cultural Heritage” of the European Fifth Framework of Research Programme) made extensive recommendations (<http://www.integaire.org/project/Integairere-searchrecommendations%5B2%5D.pdf>) still valid

today as a way forward. These recommendations, along with those addressed with the article are required future directions on the development of Urban Air Quality science.

Integaire's recommendations have been summarized below, combining them into fewer points where convenient.

1. Quantify sources of particulate matter for example resuspension of dust may contribute significantly, but it cannot be quantified.
2. Investigate the health effects of particulate matter and methods to demonstrate how PM relates to citizens' health should be developed. Especially PM_{2.5} should be monitored.
3. Assess the impact of local policies on climate change and vice versa. Some local air quality measures, in particular many emission reduction measures, are known to reduce the greenhouse effect too, but there are also local measures to reduce air pollution that have a negative effect on greenhouse gas emissions – these dynamics must be investigated.
4. Investigate the possibilities of cleaner vehicles and alternative fuel.
5. Quantify the emissions of traffic as a function of engine conditions and traffic behavior.
6. Study possibilities for controlling particulate matter and NO₂.
7. Improve the insight in the effect of buildings on hot spot concentrations.
8. Improve the possibilities of improving horizontal integration of air, water, soil, and waste policies. Strategies across these management areas must be harmonized.
9. Continue to improve insight into photochemical air pollution.
10. Develop more knowledge on the reliability of air pollution, air quality as well as air transport and deposition models, including validation procedures. Urban air quality models need to be developed and operationalized with the manager in mind, that is, results developed with the regulator in mind, integrating measurements and model results. Tools for valuation need to continue to be developed and integrated into the models.

11. Clarify and give more guidance on PM₁₀ and PM_{2.5} monitoring.
12. Develop methods for taking “soft” aspects (communication, attitudes, and participation as well as management governance structures) into account.

Given the number and complexity of the above challenges, it is also important that educational programmes and efforts to develop the air quality scientists and managers of tomorrow are increased.

There is an urgent need for field studies dedicated to a better characterization of urban particle pollution together with studies focusing at gaseous and particulate PAHs and trace elements such as heavy metals. Such studies are needed for the full assessment of the health impact on the urban population and for providing the necessary basis for future urban air pollution management. Of the pollutants dominating urban air pollution climates, particulate pollution in general together with gaseous and particulate polycyclic aromatic hydrocarbons (PAHs) and heavy metals are those where further field measurements, characterization, and laboratory studies are urgently needed in order to fully assess the health impact on the urban population and provide the right basis for future urban air pollution management.

PAHs in the urban atmosphere as well as their transformation, transport, and conversion are urgently needed to be studied. Although air quality of megacities and urban areas in most cases fall within the jurisdiction of local governments, the studies provide compelling evidence that effort at international levels to regulate LRTAP must specifically look at PAHs and their precursors. This will require an international, strategic joint collaboration among Asia, Europe, and North America to investigate these complicated mechanisms and analyze the data.

With respect to air quality models, the authors of a regional model for atmospheric photochemistry and particulate matter applied to predict the fate and transport of lead, manganese, total chromium, nickel, and cadmium, over the continental United States during January and July 2001 [63] recommend research to improve the model results on emission data for aerial suspension of particles and biomass burning. They further cite the need to better quantify anthropogenic emissions of chromium, nickel, and cadmium [in the

US National Emission Inventory (NEI)]. It is our opinion that their recommendations are valid for most NEI, which are necessary management tools for air quality policy development.

Bibliography

Primary Literature

- Bell ML, Davis DL, Fletcher T (2003) *Environ Health Perspect* 112:6–8
- WHO (2006) WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide – summary of risk assessment. WHO, Geneva
- UNFPA (2007) UNFPA State of World Population 2007 – unleashing the potential of urban growth. UNFPA, Geneva
- Hertel O, de Leeuw FAAM, Raaschou-Nielsen O, Jensen SS, Gee D, Herbarth O, Pryor S, Palmgren F, Olsen E (2001) *Pure Appl Chem* 73:933–958
- Hylander LD, Goodsite ME (2006) *Sci Total Environ* 368:352–370
- Berkowicz R, Kunzli F, Kaiser R, Medina S, Studnika M, Chanel O, Filliger P, Herry M, Horak F, Puybonnieux-Textier V, Quenel P, Schneider J, Reethaler R, Vergnaud JC, Sommer H (2000) *Environ Monit Assess* 65:259–267
- Kunzli N et al (2000) Public-health impact of outdoor and traffic-related air pollution: a European assessment. *Lancet* 356:795–801
- Gros V, Sciare J, Yu T (2007) *CR Geosci* 339:764–774
- Smith KK et al (2004) Indoor air pollution from household use of fossil fuels. WHO, Geneva
- Gurjar BR, Lelieveld J (2005) *Atmos Environ* 39:391–393
- Chan CK, Yao X (2008) *Atmos Environ* 42:1–42
- Gurjar BR, Butler TM, Lawrence MG, Lelieveld J (2008) *Atmos Environ* 42:1593–1606
- Crecelius HJ, Sommerfeld M (2005) *Gefahrstoffe Reinhalt Luft* 65:49–54
- Suppan P, Graf J (2000) *Int J Environ Poll* 14:375–381
- Farias F, ApSimon H (2006) *Environ Model Softw* 21:477–485
- Corbett JJ et al (2007) *Environ Sci Technol* 41:8512–8518
- Marr IL, Rosser DP, Meneses CA (2007) *Atmos Environ* 41:6379–6395
- Wang HB, Shooter D (2002) *Atmos Environ* 36:3519–3529
- Hedberg E, Johansson C (2006) *J Air Waste Manag Assoc* 56:1669–1678
- Sexton K, Liu KS, Hayward SB, Spengler JD (1985) *Atmos Environ* 19:1225–1236
- Maykut NN, Lewtas J, Kim E, Larson TV (2003) *Environ Sci Technol* 37:5135–5142
- Glasius M, Ketzel M, Wählin P, Bossi R, Stubkjær J, Hertel O, Palmgren F (2008) *Atmos Environ* 42:8686–8697
- Vignati E, Berkowicz R, Hertel O (1996) *Sci Total Environ* 190:467–473
- Eliasson I, Holmer B (1990) *Theo App Climatol* 42:187–196
- Rigby M, Toumi R (2008) *Atmos Environ* 42:4932–4947
- Berkowicz R, Palmgren F, Hertel O, Vignati E (1996) *Sci Total Environ* 190:259–265
- Palmgren F, Berkowicz R, Hertel O, Vignati E (1996) *Sci Total Environ* 190:409–415
- Brauer M, Hoek G, van Vliet P, Meliefste K, Fischer PH, Wijga A, Koopman LP, Neijens HJ, Gerritsen J, Kerkhof M, Heinrich J, Bellander T, Kronekreef B (2002) *Am J Respir Critic Care Med* 166:1092–1098
- Carslaw DC, Beevers SD (2005) *Atmos Environ* 39:167–177
- Beddows DCS, Beddows MCS, Donovan RJ, Harrison RM, Kinnersley RP, King MD, Nicholson DH, Thompson KC (2004) *J Environ Monit* 6:124–133
- Harrison RM, Jones AM, Lawrence RG (2004) *Atmos Environ* 38:4531–4538
- Wählin P, Berkowicz R, Palmgren F (2006) *Atmos Environ* 40:2151–2159
- Salcedo D, Onasch TB, Dzepina K, Canagaratna MR, Zhang Q, Huffman JA, DeCarlo PF, Jayne JT, Mortimer P, Worsnop DR, Kolb CE, Johnson KS, Zuberi B, Marr LC, Volkamer R, Molina LT, Molina MJ, Cardenas B, Bernabe RM, Marques C, Caffney JS, Marley NA, Laskin A, Shutthanandan V, Xie Y, Brune W, Lsher R, Shirley T, Jimenez JL (2006) *Atmos Chem Phys* 6:925–946
- Viana M, Kuhlbusch TAJ, Querol X, Alastuey A, Harrison RM, Kopke PK, Winiwarter W, Vallius A, Szidat S, Prevot ASH, Heuglin C, Bloemen H, Wählin P, Vecchi R, Miranda AI, Kasper-Giebl A, Maenhaut W, Hitzengerger R (2008) *J Aerosol Sci* 39:827–849
- Harrison RM, Shi JP, Xi SH, Khan A, Mark D, Kinnersley R, Yin JX (2000) *Philos Trans R Soc Lond A Math Phys Eng Sci* 358:2567–2579
- Löndahl J, Pagels J, Swietlicki E, Zhou JC, Ketzel M, Massling A, Bohgard M (2006) *J Aerosol Sci* 37:1152–1163
- Harrison RM, Yin JX (2008) *Atmos Environ* 42:1413–1423
- Peters A, Wichmann HE (2001) *Epidemiol* 12:S97
- Harrison RM, Yin JX (2000) *Sci Total Environ* 249:85–101
- Ketzel M, Wählin P, Kristensson A, Swietlicki E, Berkowicz R, Nielsen OJ, Palmgren F (2004) *Atmos Chem Phys* 4:281–292
- Harrison RM, Tilling R, Romero MSC, Harrad S, Jarvis K (2003) *Atmos Environ* 37:2391–2402
- Berkowicz R (2000) *Environ Monit Assess* 65:323–331
- Ketzel M, Berkowicz R (2004) Modelling the fate of ultrafine particles from exhaust pipe to rural background: an analysis of time scales for dilution, coagulation and deposition. *Atmos Environ* 38:2639–2652
- Ketzel M, Omstedt G, Johansson C, During I, Pohjolar M, Oettl D, Gidhagen L, Wahlin P, Lohmeyer A, Haakana M, Berkowicz R (2007) Estimation and validation of PM_{2.5}/PM₁₀ exhaust and non-exhaust emission factors for practical street pollution modelling. *Atmos Environ* 41:9370–9385
- Baek BH, Aneja VP (2004) *J Air Waste Manag Assoc* 54:623–633
- Pakkanen TA, Kerminen VM, Hillamo RE, Makinen M, Makela T, Virkkula A (1996) *J Atmos Chem* 24:189–205
- Pakkanen TA (1996) *Atmos Environ* 30:2475–2482

48. Schneider T, Jensen KA, Clausen PA, Afshari A, Gunnarsen L, Wählin P, Glasius M, Palmgren F, Nielsen OJ, Fogh CL (2004) *Atmos Environ* 38:6349–6359
49. International Agency for Research in Cancer (1983) IARC monographs on the evaluation of the carcinogenic risk of chemicals to humans. WHO, International Agency for Research in Cancer, Geneva
50. ATSDR (1995) Toxicological profile for polycyclic aromatic hydrocarbons (PAHs). Department of Health and Human Services, Public Health Service, Atlanta
51. Hanigan IC, Johnston FH, Morgan GG (2008) *Environ Health* 7:42
52. Denissenko MF, Pao A, Tang MS, Pfeifer GP (1996) *Science* 274:430–432
53. Molina LT, Kolb CE, de Foy B, Lamb BK, Brune WH, Jimenez JL, Ramos-Villegas R, Sarmiento J, Paramo-Figueroa VH, Cardenas B, Gutierrez-Avedoy V, Molina MJ (2007) *Atmos Chem Phys* 7:2447–2473
54. Ravindra K, Sokhi R, Van Grieken R (2008) *Atmos Environ* 42:2895–2921
55. Perwak J, Byrne M, Coons S, Goyer M, Harris J, Cruse P, Derosier R, Moss K, Wendt S (1982) Exposure and risk assessment for Benzo(a)pyrene and other polycyclic aromatic hydrocarbons, vol 4. US Environmental Protection Agency Office of Water Regulations and Standards, Washington, DC
56. Hazardous Substances Data Bank (HSDB) (1994) National Library of Medicine, National Toxicology Program (via TOXNET) <http://www.nlm.nih.gov/pubs/factsheets/hsdbfs.html>, specifically as linked in the previous to <http://toxnet.nlm.nih.gov/>. Accessed 3 February 2011
57. Marr LC, Dzepina K, Jimenez JL, Reisen F, Bethel HL, Arey J, Gaffney JS, Marley NA, Molina LT, Molina MJ (2006) *Atmos Chem Phys* 6:1733–1745
58. Dzepina K, Arey J, Marr LC, Worsnop DR, Salcedo D, Zhang Q, Onasch TB, Molina LT, Molina MJ, Jimenez JL (2007) *Int J Mass Spectrom* 263:152–170
59. Jang C, Lee J (2005) *Proc Inst Mech Eng D J Automobile Eng* 219:825–831
60. Lee JY, Kim YP (2007) *Atmos Chem Phys* 7:3587–3596
61. Kemp K (2002) *Nucl Instrum Meth Phys Res Sec B Beam Interact Mater Atoms* 189:227–232
62. Rajsic S, Mijic Z, Tasic M, Radenkovic M, Joksic J (2008) *Environ Chem Lett* 6:95–100
63. Hutzell WT, Luecken DJ (2008) *Sci Total Environ* 396:164–179
64. Hertel O, Ellermann T, Palmgren F, Berkowicz R, Løfstrøm P, Frohn LM, Geels C, Skjøth CA, Brandt J, Christensen J, Kemp K, Ketzel M (2007) *Environ Chem* 4:65–74
65. Hertel O, Jensen SS, Hvidberg M, Ketzel M, Berkowicz R, Palmgren F, Wählin P, Glasius M, Loft S, Vinzents P, Raaschou-Nielsen O, Sørensen M, Bak H (2007) Assessing the impact of traffic air pollution on human exposures and linking exposures to health effects. In: Jensen-Butler C, Sloth B, Larsen MM, Madsen B, Nielsen OA (eds). *Road Pricing, the Economy and the Environment*, Series: Advances in Spatial Science, Springer, Berlin, p 425
66. Hertel O, Goodsite ME (2009) Urban air pollution climates throughout the world. In: Hester RE, Harrison RM (eds) *Air quality in urban environments*, vol 28. RSC, Cambridge, UK, pp 1–22

Books and Reviews

- Anderson LG (2009) Ethanol fuel use in Brazil: air quality impacts. *Energy Environ Sci* 2(10):1015–1037
- Chan CK, Yao X (2008) Air pollution in mega cities in China. *Atmos Environ* 42(1):1–42
- Cimino-Reale G et al (2008) Toxicogenomic study of indoor and outdoor air chemical mixtures. *Fresenius Environ Bull* 17.9B:1485–1491
- Colbeck I, Nasir ZA, Ali Z (2010) The state of ambient air quality in Pakistan—a review. *Environ Sci Pollut Res* 17(1):49–63
- Deguen S, Zmirou-Navier D (2010) Social inequalities resulting from health risks related to ambient air quality—A European review. *Eur J Public Health* 20(1):27–35
- Ebi KL, McGregor G (2008) Climate change, tropospheric ozone and particulate matter, and health impacts. *Environ Health Perspect* 116(11):1449–1455
- El-Fadel M, bi-Esber L (2009) In-vehicle exposure to carbon monoxide emissions from vehicular exhaust: a critical review. *Critic Rev Environ Sci Technol* 39(8):585–621
- Farmaki EG, Thomaidis NS (2008) Current status of the metal pollution of the environment of Greece – a review. *Glob Nest J* 10(3):366–375
- Goyal SK et al (2006) Understanding urban vehicular pollution problem vis-a-vis ambient air quality – case study of a megacity (Delhi, India). *Environ Monit Assess* 119(1–3):557–569
- Gros V, Sciare J, Yu T (2007) Air-quality measurements in megacities: focus on gaseous organic and particulate pollutants and comparison between two contrasted cities, Paris and Beijing. *CR Geosci* 339(11–12):764–774
- Hidalgo J et al (2008) Advances in urban climate modeling. *Trends Dir Climate Res* 1146:354–374
- Holmes NS, Morawska L (2006) A review of dispersion modelling and its application to the dispersion of particles: an overview of different dispersion models available. *Atmos Environ* 40(30):5902–5928
- Jacob DJ, Winner DA (2009) Effect of climate change on air quality. *Atmos Environ* 43(1):51–63
- Just J et al (2007) Air pollution and asthma in children. *Rev Fr D Allergol D Immunol Clin* 47(3):207–213
- Kanda M (2007) Progress in urban meteorology: a review. *J Meteorol Soc Jpn* 85B:363–383
- Kindzierski WB et al (2009) Automotive wastes. *Water Environ Res* 81(10):1545–1568
- Krewski D, Rainham D (2007) Ambient air pollution and population health: overview. *J Toxicol Environ Health A* 70(3–4):275–283

- Kuykendall JR et al (2009) Chemicals present in automobile traffic tunnels and the possible community health hazards: a review of the literature. *Inhal Toxicol* 21(8–11):747–792
- Liu Y et al (2009) Review of the applications of multiangle imaging spectroradiometer to air quality research. *Sci China, Ser D Earth Sci* 52(1):132–144
- Mejia JF, Morawska L (2009) An investigation of nucleation events in a coastal urban environment in the Southern Hemisphere. *Atmos Chem Phys* 9(20):7877–7888
- Morawska L et al (2008) Ambient nano and ultrafine particles from motor vehicle emissions: characteristics, ambient processing and implications on human exposure. *Atmos Environ* 42(35):8113–8138
- Nuvolone D, Barchielli A, Forastiere F (2009) Assessing the effectiveness of local transport policies for improvements in urban air quality and public health: a review of scientific literature. *Epidemiol Prev* 33(3):79–87
- Ooi GL (2007) Urbanization in Southeast Asia – assessing policy process and progress toward sustainability. *J Ind Ecol* 11(2):31–42
- Paoletti E (2009) Ozone and urban forests in Italy. *Environ Pollut* 157(5):1506–1512
- Ravindra K, Sokhi R, Van Grieken R (2008) Atmospheric polycyclic aromatic hydrocarbons: source attribution, emission factors and regulation. *Atmos Environ* 42(13):2895–2921
- Rodriguez S et al (2007) Atmospheric particulate matter and air quality in the Mediterranean: a review. *Environ Chem Lett* 5(1):1–7
- Sarnat JA, Holguin F (2007) Asthma and air quality. *Curr Opin Pulm Med* 13(1):63–66
- Sidiropoulos C, Tsiliniridis G (2009) Improved sectoral allocation of NMVOC emissions from solvent use in Greece. *Sci Total Environ* 407(13):4075–4083
- Sitaras LE, Siskos PA (2008) The role of primary and secondary air pollutants in atmospheric pollution: Athens urban area as a case study. *Environ Chem Lett* 6(2):59–69
- van den Elshout S, Leger K, Nussio F (2008) Comparing urban air quality in Europe in real time – a review of existing air quality indices and the proposal of a common alternative. *Environ Int* 34(5):720–726
- Vardoulakis S et al (2007) Operational air pollution modelling in the UK – street canyon applications and challenges. *Atmos Environ* 41(22):4622–4637
- Vautard R et al (2007) Air quality in Europe during the summer of 2003 as a prototype of air quality in a warmer climate. *CR Geosci* 339(11–12):747–763
- Whitehead PG et al (2009) A review of the potential impacts of climate change on surface water quality. *Hydrol Sci J J Sci Hydrol* 54(1):101–123
- Yearley S (2006) Bridging the science – policy divide in urban air-quality management: evaluating ways to make models more robust through public engagement. *Environ Plann C Gov Policy* 24(5):701–714
- Zhang JJ, Smith KR (2007) Household air pollution from coal and biomass fuels in China: measurements, health impacts, and interventions. *Environ Health Perspect* 115(6):848–855

Urban Atmospheric Composition Processes

WILLIAM BLOSS

School of Geography, Earth & Environmental Sciences,
University of Birmingham, Birmingham, UK

Article Outline

Glossary
Definition of the Subject
Introduction
Historical Perspective
Meteorology, Dynamics and Mixing
Basic Composition and Key Pollutants
Key Emission Characteristics
Urban Atmospheric Composition Processes
Trends in Urban Atmospheric Composition
Future Directions
Bibliography

Glossary

Boundary layer The lowermost layer of the atmosphere, in contact with the surface and well mixed. Extending from 100 to 200 m at night, up to 1–2 km during the day.

Mixing ratio The fraction of the air which a given species comprises, typically parts per million (10^6), billion (10^9) or trillion (10^{12}), by volume.

NO_x The sum of nitric oxide (NO) and nitrogen dioxide (NO₂).

NO_x limited The ozone production regime in which VOCs are (relatively) plentiful, and ozone production is limited by the availability of NO_x.

Photostationary steady state The rapid chemical/photochemical balance between a number of reactants, usually applied to NO, NO₂ and O₃.

PM₁₀ Suspended airborne particulate matter, with an aerodynamic diameter of 10 µm or less.

Primary pollutants Those emitted directly to the atmosphere, typically from mechanical or combustion processes.

Secondary pollutants Those formed within the atmosphere by chemical reactions.

Street canyon The semi-isolated atmospheric compartment formed by adjacent rows of multi-story buildings.

Temperature inversion An increase in temperature with increasing altitude, which usually isolates ground-level air from the overlying atmosphere, resulting in increased (primary) pollutant concentrations.

Transboundary pollution The transport of long-lived chemical species, such as O_3 and CO , across national boundaries.

VOC Volatile Organic Compound, one commonly found overwhelmingly in the gas phase.

VOC limited The ozone production regime where NO_x is plentiful, and ozone production is limited by the availability of hydrocarbons.

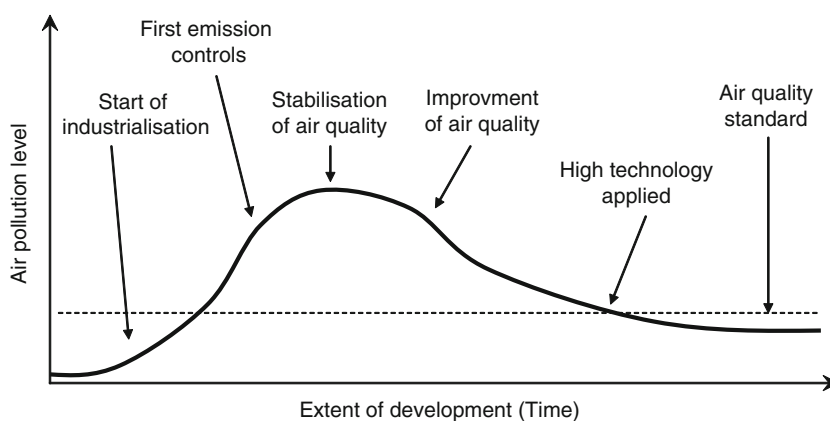
Definition of the Subject

The majority of the population now live in cities (the 50% point was passed during 2007 [1]) and are exposed to the urban atmosphere every day of their lives. As global population has also doubled over the past 50 years, urban air quality is a major factor in pollutant exposure for billions of individuals. The urban atmospheric environment is unique, with the city landscape defining particular meteorological characteristics, and a wide range and high density of emission (or pollution) sources, which combine to determine local air quality. Air pollutants, ranging from gases such as carbon monoxide, ozone and

nitrogen dioxide, to suspended particles of varying size and composition, have been shown to be linked to respiratory and cardiovascular health issues. Urban atmospheric composition processes affect the abundance of primary pollutants, and are wholly responsible for the formation of secondary species such as ozone. In this article, the principle urban atmospheric components and their sources are introduced, followed by an account of the predominant atmospheric composition (chemistry) processes which affect urban air quality. The article concludes with a brief consideration of the potential impacts of future changes in emissions, physical and political climates upon urban atmospheric composition.

Introduction

Increasing population density implies an increasing human exposure to urban atmospheric pollutants. Countering this basic trend, growing scientific understanding of atmospheric composition processes, coupled with technological developments in emissions control and policy measures to limit pollution, have led to improved urban air quality over the past several decades in many cities, although this conclusion is dependent upon the species measured and metric chosen. The predominant air quality conditions are therefore a strong function of the level of development of a city (Fig. 1), and have changed substantially over the past few decades, with (in many cases) the removal of heavy industry from urban regions, while traffic has



Urban Atmospheric Composition Processes. Figure 1

Schematic representation of the evolution of air pollution vs. development. Adapted from Fenger [2]/Mage et al. [3]

become the dominant emission source; the concept of urban atmospheric pollution thus varies in time, and between different cities across the world.

Historical Perspective

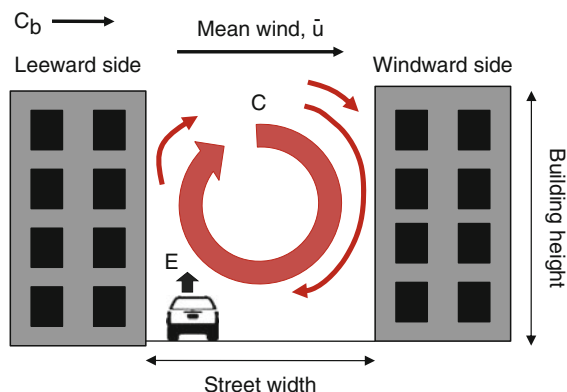
The history of urban atmospheric composition and air pollution reflects developments in fuel usage for domestic heating, cooking, emerging industrial activities and power generation, and the spread of the internal combustion engine. The earliest reports of atmospheric pollution issues, associated with the combustion of coal for domestic heating and cooking, and for industrial operations such as lime production, are noted in historical documents dating to at least the thirteenth century, when a Commission was established to address the problems associated with coal smoke in London. By the sixteenth or seventeenth century, some of the earliest proposals for “emission control legislation” such as John Evelyn’s *Fumifugium* of 1661 were drafted for (unsuccessful) presentation to Parliament [4]. Increasing domestic and industrial coal combustion led to the London smogs of the nineteenth and twentieth centuries, in which the major pollutants were highly acidic particulates and fog droplets arising from use of high-sulfur fuels. Another prominent atmospheric pollution event of the time was the Donora Valley episode of 1948, in which local industrial emissions of sulfur dioxide, carbon monoxide and heavy metal particles were trapped in a week-long temperature inversion leading to tens of deaths and thousands suffering from respiratory illness. These events primarily concerned primary pollutants, those emitted directly to the atmosphere (although strictly speaking sulfuric acid is formed within the atmosphere following the release of sulfur dioxide). Subsequent legislation, mandating the use of tall chimneys to aid dispersion, use of low-sulfur and smokeless fuels in populated areas, the migration of heavy industry away from cities and the introduction of emissions control technologies and policies (including emissions trading schemes on the case of SO_2), have greatly reduced these pollution phenomena in many nations. Over the past 50 years, motor vehicle emissions have come to dominate urban atmospheric composition in most developed nations, as the global vehicle fleet has increased by an order of magnitude over this timeframe.

In developing nations, combustion of biofuels (primarily wood) and coal for domestic cooking and heating remain major urban pollutant sources. Alongside the reduction in some of the traditional primary pollutants (sulfur dioxide, carbon monoxide), levels of secondary components formed in the atmosphere, such as ozone and (much of) nitrogen dioxide and particulate matter, have in many cases been rising, and have become the focus of concern.

Meteorology, Dynamics, and Mixing

The urban atmosphere may be defined horizontally by the extent of the particular city in question (see comments below regarding export of urban atmospheric pollutants), and vertically as encompassing the boundary layer, the first few 100 m to kilometer or so of the atmosphere, which is regarded as well mixed. The urban boundary layer height determines the size of the atmospheric compartment into which emissions are mixed – a lower boundary layer height (as commonly found at night) results in higher concentrations for an equivalent emission strength, all other things being equal. The urban boundary layer may be further divided into vertical regions determined by the interaction between the city topography and airflow: At the street scale, urban canyons, formed between rows of buildings, represent semi-isolated environments into which traffic-related emissions are initially discharged. The urban canopy layer encompasses multiple street canyons at rooftop level. Above the canopy layer, the overlying boundary layer is sometimes divided into the roughness sub-layer, where building wake-induced turbulence effects predominate and atmospheric composition (as far as it relates to urban emissions) is highly variable, and the higher inertial sub-layer which exhibits a more homogeneous composition [5].

Local wind speed and direction critically affect local pollutant levels – on a street canyon scale, winds perpendicular to the canyon may lead to the establishment of within-canyon vortices, leading to the upwind side of the canyon, at ground level, experiencing higher levels of primary pollutants (those from local traffic at least) – Fig. 2. Conversely, along-canyon wind effectively ventilates the ground-level emissions into the overlying boundary layer. While most pollutant levels fall with increasing windspeed as a consequence of



Urban Atmospheric Composition Processes. Figure 2
Typical within-canyon vortex circulation, driven by overlying perpendicular airflow, increases primary pollutant levels on the upwind (leeward) side of the street at ground level

increased dilution, re-suspension of coarse particulate matter (dust, sand, tyre, and brake debris) is found to increase atmospheric PM (particulate matter) levels at elevated windspeeds [6].

The ventilation of the urban airshed as a whole affects the pollutant levels experienced for a given emission source. Certain cities (notable examples include Mexico City, Santiago, Los Angeles) are partially to almost entirely surrounded by mountain ranges, leading to build-up of emissions within the (urban) boundary layer (e.g., Molina and Molina [7]). In other locations, regional weather patterns lead to persistent patterns of pollutant exposure, for example, the daily sea breeze circulation affecting Hong Kong and the Pearl River Delta [8]. Local circulation patterns and topography can therefore exacerbate, or ameliorate, the atmospheric composition resulting from a given emissions situation [9]; and the combined effects of local and regional scale dynamics, mixing and chemical transformations must be considered in the development of urban air quality management policies.

Basic Composition and Key Pollutants

The bulk atmospheric composition is essentially constant throughout the troposphere, comprising gases such as nitrogen, oxygen, argon, helium, carbon dioxide and a variable amount of water vapor. Most concern over urban atmospheric composition and air

pollution arises from the trace components – gaseous constituents such as ozone, nitrogen oxides and volatile organic compounds, and suspended particulate material (PM) commonly referred to as aerosol (although strictly the term denotes both the condensed phase material and the gas it is suspended within).

Urban air pollution issues primarily relate to local emissions and the chemical processing of pollutants and natural atmospheric species which occurs on time-scales of a few hours to a few days. This can be distinguished from the wider, regional impacts of air pollution (such as increasing background ozone levels, and the abundance of sulfate aerosol), although these also impact on the urban environment and are considered here. Global air pollution issues related to long-lived species such as CFC, HCFC and halon compounds contributing to stratospheric ozone depletion, and the climate-related issues associated with the global abundance of longer-lived greenhouse gases such as carbon dioxide and methane, are not discussed further here – although some urban air pollutants are also important greenhouse gases, for example, ozone. Table 1 gives typical abundances for key urban air components which would be found in the remote boundary layer, away from anthropogenic emission sources, and compares these with those observed in selected urban atmospheres. It should be noted that point measurements such as these may be representative of a local microenvironment only, rather than the wider city-scale environment – if cities do not have representative networks for air quality monitoring, this concern persists into reported pollution levels.

As understanding of the impact of air pollutants upon health has developed, various national and international air quality standards and targets have been established, by national governments, the European Union (EU), and the World Health Organization (WHO). Selected current standards for key urban pollutants are summarized in Table 2. These species – the gases nitrogen dioxide NO_2 , ozone (O_3), carbon monoxide (CO), sulfur dioxide (SO_2), benzene (C_6H_6), 1,3-butadiene (C_4H_6), and polycyclic aromatic hydrocarbons (PAH), plus the integrated particulate matter (PM) measures PM_{10} and $\text{PM}_{2.5}$ (see below for definitions) are the focus of this chapter. An issue in comparing pollution levels is the differing units commonly used – measurements of trace gases for scientific

Urban Atmospheric Composition Processes. Table 1 Comparison of typical unpolluted boundary layer abundance of selected pollutants, and their equivalent levels in various urban centers. All concentrations expressed in $\mu\text{g m}^{-3}$

Species	Remote MBL	London: MR	London: NK	Beijing	Lagos
NO ₂	0.04	98	37	66	95
O ₃	58	18	34	109 ^b	64
SO ₂	0.2	7.7	1.9	53	186
CO	140	650	280	2,100	6,700
Benzene	0.01	1.3	–	13.4 ^a	–
1,3-butadiene	<0.01	0.57	–	–	–
PM ₁₀	–	27	17	161	228
PM _{2.5}	–	18	11	130 ^b	–

Notes: Remote MBL (Marine Boundary Layer) data taken from the Cape Verde Atmospheric Observatory in the tropical Atlantic Ocean [10, 11] except SO₂, benzene, 1,3-butadiene from Mace Head. London NK (North Kensington) is an urban background site while London MR (Marylebone Road) is a roadside monitoring station on a congested major road. London data are annual averages (for 2010) taken from the UK Air Quality Archive (www.airquality.co.uk). PM data are TEOM FDMS measurements. Beijing data are averages across the city for 2006, as given by [12] except (a): [13] and (b) [14]. Lagos data are averages of several sites taken from [15] except O₃ from [16]

Urban Atmospheric Composition Processes. Table 2 (Selected) EU air quality limit values and WHO air quality guidelines. Values in $\mu\text{g m}^{-3}$

Species	EU	WHO
NO ₂	40 (annual mean)	40 (annual mean)
O ₃	120 (target, <25 times/year)	100 (8 h mean)
SO ₂	124 (24 h mean, <3 times/year)	20 (24 h mean)
CO	10,000 (running 8 h mean)	10,000 (8 h mean)
Benzene	5 (annual mean)	–
1,3-butadiene	2.25 (UK, annual mean)	–
PM ₁₀	40 (annual mean)	20 (annual mean)
PM _{2.5}	25 (2020 target)	10 (annual mean)

EU limits are those in force as of 2010; *target* values not legally binding. Additional limit values/future reductions also apply for some pollutants. WHO guidelines from 2005 update [17] except CO (1997 update)

purposes are frequently presented in terms of (volume) mixing ratio (vmr), in parts-per-million, billion or trillion (10^6 , 10^9 , and 10^{12} , respectively) while for PM, and for legislative purposes, concentration values (usually expressed as $\mu\text{g m}^{-3}$) are used. Conversion between these is straightforward but is also dependent upon the local temperature and pressure; care is needed

as different standard conditions for the conversion are mandated by, for example, the EU (20°C) and the WHO (25°C). While mixing ratio units are commonly preferred for atmospheric chemistry research where gaseous species are concerned, as they are invariant to changes in temperature and pressure (i.e., air parcel transport), and preserve the stoichiometry of the

chemical reactions (i.e., the relationship between reactants and products), here mass concentration units are used for consistency and ease of comparison with air quality standards.

Key Emission Characteristics

Ambient urban air pollutant emissions in most developed regions are primarily associated with vehicle traffic – direct exhaust emissions from petrol (spark) and diesel (compression) powered vehicles, and mechanical sources such as brake and tyre erosion and road surface material and dust re-suspension from all vehicles. Fugitive emissions of fuel vapor from filling station operations are a further significant vehicle-related source of gaseous organic (hydrocarbon) air pollutants. Other significant sources include emissions from local point sources such as domestic and office heating, cooking, and industrial processes, in particular the use of solvents, paints and lubricants which may have a significant vapor pressure, and combustion processes. Large-scale industrial emissions associated with heavy industry (power generation, steel making, etc.) are not major contributors to general *urban* air quality issues in most developed nations, as the plants responsible have largely been removed from the city centers, and as readily identifiable point sources in most cases have their direct emissions scrubbed at source, for example, in the case of sulfur dioxide from coal-fired power stations. In urban regions within developing nations, which incorporate many of the fastest-growing megacities on the planet, a different picture of emission sources is observed; again vehicle emissions are a major contributor, although the fleet composition differs with typically older vehicles (i.e., conforming to less stringent, if any, emission regulations), reduced diesel prevalence and a greater contribution from small two-stroke (oil burning) powered two wheelers. Domestic heating and cooking, where this is primarily through wood burning, is a major contributor to urban atmospheric composition (and indeed to overall pollutant exposure, through indoor effects within the home).

Characteristics of Motor Vehicle and Domestic Fuel Emissions

Within internal combustion engine vehicles, emissions of a range of pollutants reflect the composition of the

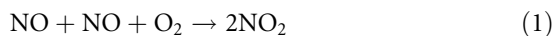
fuel and oil burnt, and the high-temperature decomposition of atmospheric oxygen and nitrogen leading to the production of nitrogen oxides. NO_x ($\text{NO} + \text{NO}_2$) at the tailpipe is conventionally assumed be partitioned (on average) into 95% NO and 5% NO_2 , although there is some evidence that the direct fraction of NO_2 may be greater than this, and increasing across the fleet on average – in particular, for diesel-engined HGVs (heavy goods vehicles) [18]. Incomplete combustion leads to the production of carbon monoxide (CO), a vast range of hydrocarbons which may be gases (if they are sufficiently small and volatile) or may condense to form the organic component of exhaust particulates, as the exhaust gases cool. The introduction of catalytic converters has greatly reduced emissions of NO , CO and hydrocarbons (which are converted to N_2 , CO_2 and H_2O) for petrol-engined vehicles, while diesel oxidation catalysts have reduced CO and hydrocarbon emissions from diesel engines. Increasingly stringent emissions regulations, such as the EURO-1 to EURO-5 EU series of emissions standards for various vehicle classes, have substantially reduced emissions per unit activity (at least within the parameters of the defined driving cycles), and (disregarding CO_2) the principal emission concerns from modern vehicles are NO_x and particulate matter, for which in both cases diesel engine emissions tend to be higher emitters. Exhaust PM in particular is known to incorporate toxic and carcinogenic components such as PAH and heavy metals, and (as new particles are formed through condensation/nucleation) to contribute substantially to the population of the smallest particles particularly associated with adverse health effects [19]. The introduction of biofuels (ethanol) in some markets (notably, South America) has been associated with increased emissions of aldehydes such as formaldehyde and acetaldehyde [20] which are toxic, and also potent radical precursors leading to photochemical smog formation (see below). Biofuel (wood and charcoal) combustion for domestic heating and cooking is a substantial source of pollutants within many homes in developing nations, leading to indoor air pollutant exposure, and forming a major emission source within urban regions. The relatively low temperatures commonly found lead to substantial emissions of carbon monoxide, black carbon (soot) and a range of trace compounds VOCs, oxygenated compounds and acids, including toxic

substances such as polycyclic aromatic hydrocarbons (PAHs), particularly within the organic component of the emitted particulate matter [21].

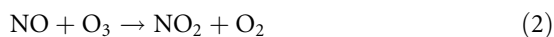
Urban Atmospheric Composition Processes

Gas-Phase Composition Processes

Gas-phase processes in urban environments are dominated by the oxides of nitrogen, (primarily) NO and NO₂, emitted directly from the exhausts of high-temperature (i.e., internal combustion engine) vehicles. Following emission, a number of gas-phase processes follow. In close proximity (within a few meters) to the tailpipe, NO levels may be sufficiently high before mixing and dilution that the NO-self-reaction, converting NO into NO₂, may occur:



However the dominant chemistry of NO_x in the urban environment is the interaction between NO, NO₂, and ozone. In daylight, these three species are very closely coupled by the NO+O₃ reaction, and the solar decomposition (photolysis) of NO₂:



The O atoms formed through NO₂ photolysis rapidly and near-exclusively recombine with molecular oxygen (O₂) to form ozone (reaction 4), where M denotes the involvement of a third body, nitrogen or oxygen, which removes some of the excess energy from the nascent ozone molecule.



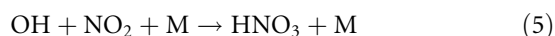
As these reactions dominate the NO/NO₂/O₃ interconversion, it is possible to define two further quantities, NO_x (= NO+NO₂), and total oxidant O_x (= NO₂+O₃), which, in the absence of further processes, are conserved. NO_x and O_x provide useful metrics for assessing variations in the sum of NO and NO₂, and of NO₂ and O₃, without need to consider the partitioning between these species. The reaction of fresh, vehicle-emitted NO with O₃ reduces ozone levels in many urban areas compared with the surrounding regions, leading to an *urban decrement* in ozone (compare O₃ concentrations for the MR and NK sites in Table 1) – discussed further below.

The timescales of reactions 2 and 3 in the boundary layer under sunlit conditions is of the order of 1–2 min, while reaction 4 proceeds in milliseconds; consequently away from direct emission sources NO, NO₂ and O₃ are found in equilibrium defined by the Leighton Relationship [22]:

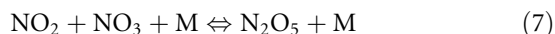
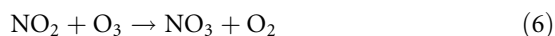
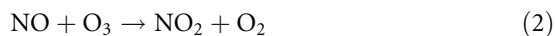
$$\frac{[\text{NO}_2]}{[\text{NO}]} = \frac{k_2[\text{O}_3]}{j_3}$$

where k_2 is the rate constant for the NO+O₃ reaction, and j_3 is the (solar radiation intensity-dependent) rate of photolysis of NO₂. It follows that for a given abundance of NO_x and O₃, in the absence of other processes, NO_x will exist primarily as NO₂ when light levels are lower (i.e., at dawn and dusk), while the NO fraction will maximize at midday. In practice the values of k_2 and j_3 are such that the NO₂:NO ratio is typically in the range 2–5 during the day – see Fig. 3.

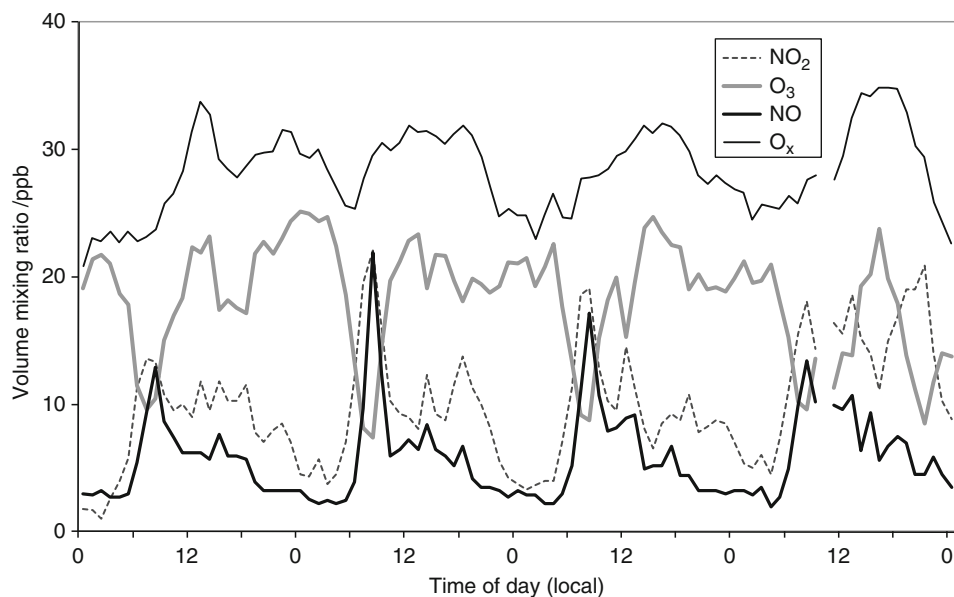
NO_x is removed from the atmospheric system through chemical conversion to more stable species, which in turn are removed by deposition and by their uptake or incorporation into aerosol particles. During the day, NO₂ reacts with the OH radical, the gaseous oxidant which dominates the removal of many pollutants on a global scale, forming nitric acid, HNO₃. Nitric acid is then lost through dry deposition (i.e., contact with surfaces), wet deposition/rainout, and through uptake into aerosol particles (see below).



The OH radical is formed primarily through sunlight-driven photolysis processes, so has a very low abundance at night; in the absence of sunlight, the NO₂ photolysis reaction also ceases, and a separate chain of reactions dominate NO_x:



The nitrate radical, NO₃, is rapidly photolysed during the day, effectively suppressing reactions 6 and 7. At night however NO and NO₂ may be converted into N₂O₅ on a timescale of hours; N₂O₅ is then removed from the urban atmosphere through wet and dry deposition, and uptake into aerosol particles (it is also photolysed slowly during the day). In both cases,

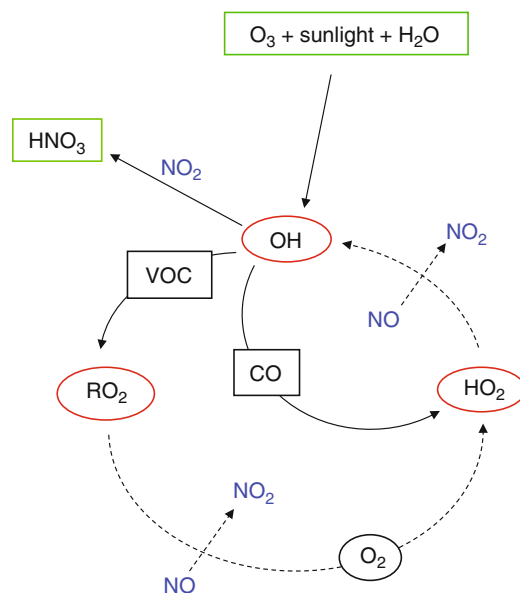


Urban Atmospheric Composition Processes. Figure 3

Diurnal variation in NO, NO₂, O₃, and O_x over a 4-day period for Birmingham City Centre; from Bloss [23]

following removal from the gas phase HNO₃ and N₂O₅ contribute to the nitrate (NO₃⁻) content of aerosol particles (see below), and to the acidity of precipitation. The combination of reactions 5 and 7 determine a chemical lifetime for NO_x (strictly, time for concentrations to fall to 1/e of their initial value) of the order of 12–24 h in urban environments.

A further major air pollution consequence of the interaction of NO_x and VOCs in sunlight is the photochemical production of ozone, a major secondary pollutant. Figure 4, below, summarizes the major reaction processes responsible (after Wayne, [24]). The process is initiated by the production of an oxidant radical, OH, shown in Fig. 4 from the photolysis of ozone in the presence of water vapor (although other indirect sources also occur, e.g., aldehyde photolysis). OH radicals react with organic compounds (VOCs and CO) which leads to the formation of organic and hydro peroxy radicals, RO₂ and HO₂. In the absence of NO₂, these undergo self-reaction to form peroxides (not shown); however in the urban environment substantial levels of NO are usually present, and the peroxy radicals react with NO, converting it to NO₂, and



Urban Atmospheric Composition Processes. Figure 4

Some important steps in the gas-phase oxidation of volatile organic compounds (VOCs) in the presence of NO_x. The photolysis of NO₂ regenerates NO and leads to net ozone production

regenerating the radical species HO_2 and subsequently OH. The peroxy radical–NO reaction therefore leads to the conversion of NO to NO_2 without the consumption of O_3 (cf. reaction 2), and the subsequent photolysis of NO_2 and reaction of the O atom formed with molecular oxygen (reactions 3 and 4) results in ozone formation. While initiation of the process consumes ozone, the cyclical nature of the reactions leads to net ozone production overall.

The timescale of ozone production ranges from hours to days, but is complex to determine directly. In the boundary layer, ozone removal is dominated by deposition, however as the atmospheric lifetime of ozone is several days to weeks, advection (of ozone-rich or ozone-poor air from upwind locations) and atmospheric mixing also have a substantial influence upon measured ozone levels. Observed ozone pollution episodes vary in extent from hours in highly polluted conditions to a few days, and are closely associated with the prevailing weather conditions. In the case of Western Europe, ozone pollution events commonly correspond to summertime anticyclonic periods in which air masses have looped over Europe accumulating pollutants. Ozone is chemically formed along the air mass trajectory at rates of the order of 20–30 ppb per day (daytime chemical production, part of which is offset by night-time chemistry and deposition [25]). If a typical episode is characterized by ozone levels of the order of 90 ppb compared with the rural background of around 30–40 ppb, several days are clearly required to achieve this level. In more extreme conditions, ozone production proceeds more rapidly: during the 2003 Western Europe heatwave event, a photochemical model constrained by in situ observations of long-lived species from suburban North-East London determined the peak ozone production to be 17 ppb h^{-1} , with an average production rate during midsummer at midday of 7.2 ppb h^{-1} [26]. Observations in Mexico City have led to calculated ozone production rates of up to 50 ppb h^{-1} [27] during stagnant meteorological conditions.

Exposure to ozone has substantial human health consequences, being associated with irritation to eyes and nose, inflammatory reactions and reduced lung

function. Ozone is also harmful to vegetation and ecosystems, and certain materials (erosion of polymers through reaction with component double bonds), and is an important greenhouse gas in its own right; consequently reduction in ozone levels is a major goal of air quality policy. The long lifetime of ozone limits the impact that local/national air quality legislation can have upon boundary layer ozone levels, and dictates that ozone is treated as a *transboundary* pollutant.

The secondary nature of ozone further complicates attempts to reduce its abundance: The precursor species, NO_x and VOCs, must be controlled; however the response of the chemical ozone production rate to the level of, for example, NO_x and CO is not straightforward: As can be seen from inspection of Fig. 3, at high NO_x levels, the reaction of OH with NO_2 , forming HNO_3 , can compete with the reaction of OH with VOCs. In this instance, reducing NO_x levels (e.g., through emission controls) would have the effect of increasing the abundance of OH, and the rate of the OH+VOC reactions, and hence would lead to an increase in ozone production. At lower NO_x levels, where NO_2 is not competing effectively with VOCs for OH, reduction in NO_x leads to a reduction in the ozone production rate, by reducing the rate of the peroxy radical–NO reaction (or strictly, by allowing other competing processes, not shown on Fig. 3, to occur – [24]). Two overall regimes for ozone production can be identified: VOC-limited, where NO_x levels are high, and ozone production rates will rise or fall with increase/decreases in VOC levels respectively, and NO_x -limited, where VOCs are relatively more abundant, NO_x levels low, and ozone levels rise and fall with increases/decreases in NO_x . Reduction of NO_x in a VOC-limited regime may lead to increased ozone production. Many urban environments correspond to VOC-limited conditions (i.e., higher NO_x levels); as air is advected from the urban centre to surrounding rural regions, with reduced emission sources, NO_x levels fall and the ozone production regime undergoes a transition to a NO_x -limited state. The complex dependence of the ozone production rate upon the abundance of primary pollutants between urban and rural areas, which is compounded by the reduction in ozone levels on the street canyon scale within

urban areas by the titration of O_3 to NO_2 by fresh vehicle NO emissions, presents a major challenge to the development of effective policy for mitigation of ozone pollution – see below.

Different organic compounds have differing potentials for ozone production (reflecting their chemical identity/size, and the rates of their reactions with oxidants such as OH), and of course are emitted in varying quantities. Metrics to rank species in terms of their contribution to ozone levels (such as Photochemical Ozone Creation Potentials, POCP values) have been developed to account for the integrated effect of these factors; in the case of the UK, the highest POCP values are assigned to tri-methylbenzene, however considering the mass-weighted emissions, the largest contributors to the ozone levels experienced are toluene and n-butane [28]. These values are specific to the UK case, that is, to the emissions, and in particular NO_x levels, typically experienced by an air parcel advected from continental Europe to southern Britain.

Peroxy acetyl nitrate (PAN, $CH_3C(O)OONO_2$) is the simplest example of a range of acetyl nitrate compounds, secondary pollutants which are the principal lachrymator (tear inducing) component of photochemical smog. PAN is formed from the reaction between peroxy acetyl radicals, $CH_3C(O)OO$, and NO_2 , and in terms of the chemical cycling acts as a temporary reservoir for the HO_x and NO_x radicals involved in ozone production. A key feature of the equilibrium between PAN and its precursor species is the strong temperature dependence, with lower temperatures favoring PAN stability. A scenario which arises then is that PAN is formed in a polluted urban area, transported aloft (where temperatures are lower and PAN is stable) over long distances, before air mass descent and warming leads to PAN decomposition and release of the component NO_x in otherwise unpolluted regions.

Sulfur dioxide (SO_2) is possibly the classic air pollutant formed from the combustion of sulfur-containing fuels. SO_2 emissions have been successfully reduced in many regions, through a combination of the use of low sulfur fuels, and application of flue gas desulphurization technologies on major plants (e.g., power stations). Within the atmosphere, SO_2 is processed to H_2SO_4 on a timescale of days, both in the gas phase, initiated by reaction with the OH radical,

and in the condensed phase, within water droplets, where the oxidation reactions are driven predominantly by hydrogen peroxide, H_2O_2 . Gas phase sulfuric acid has a very low vapor pressure, and is readily incorporated into aerosol and condensation, raising PM acidity. More widely, co-condensation of H_2SO_4 with water vapor and/or ammonia leads to the production of new particles in the atmosphere.

Condensed Phase Composition Processes

Particulate matter is an important component of urban atmospheric composition, in particular where health effects are concerned. Sources of particulate matter include the local primary emissions from vehicles and combustion described above, including mechanical sources related to vehicle brake pad/disc, tyre and road surface wear, wider atmospheric inputs such as wind-blown dust and crustal material, and secondary contributions arising from the condensation of low-volatility gases to form new or (predominantly) contribute to the growth of existing particles.

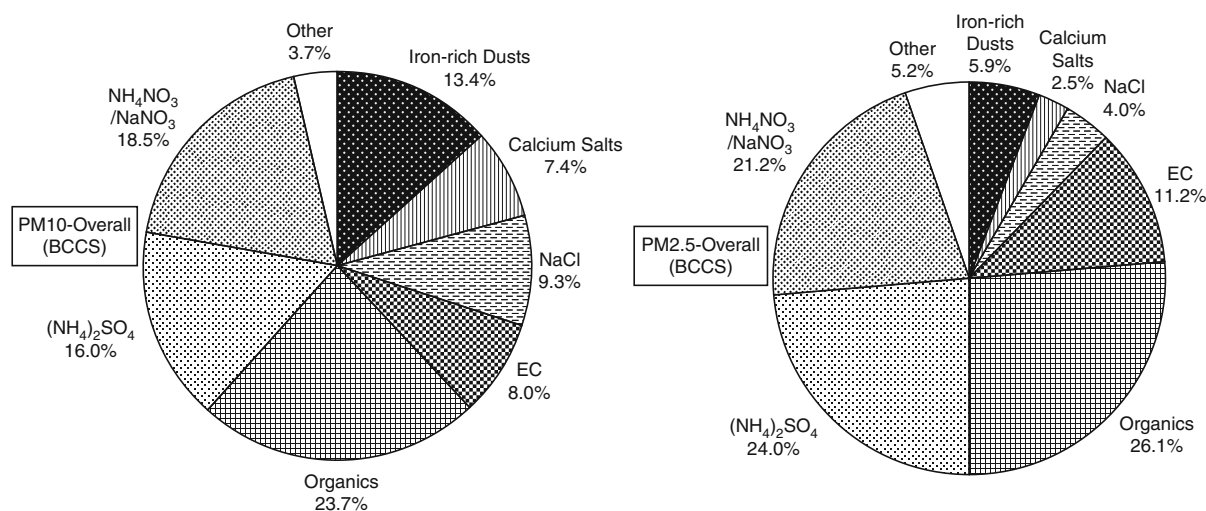
Levels of particulate matter or aerosol are usually expressed as a mass concentration below a certain size; the size ranges commonly selected relating loosely to the extent to which particles penetrate the human respiratory system. PM_{10} and $PM_{2.5}$ define the fractions of suspended particulate matter with (aerodynamic) diameters below 10 and 2.5 μm respectively (the aerodynamic diameter is a concept used to allow for the varying morphology or shape of particles; a particle with an aerodynamic diameter of 1 μm will exhibit the same inertial properties as a sphere with a diameter of 1 μm and a density of 1 $g\ cm^{-3}$ – irrespective of the actual size, shape or density of the particle). PM_{10} and $PM_{2.5}$ correspond approximately to size ranges which may be inhaled into the thoracic cavity (lungs) and into the alveolar regions respectively. The size distribution of urban particles typically exhibits several modes, reflecting the varying dominance of different processes responsible for particle formation and growth: The nucleation mode, below approximately 0.1 μm in diameter, where secondary particles are formed from the condensation of low-volatility gases (or, more commonly, condenses onto pre-existing particles); the accumulation mode (0.1–2.5 μm), where particles grow by condensation

and coagulation (collisions), but within which the mass is too small for gravitational settling to be a major sink (hence the size range where particles accumulate, until removal primarily by precipitation) and the coarse mode, which incorporates most mechanically generated particles, and within which gravitation settling begins to be a significant loss process. As simple geometric considerations show, the largest number of particles is found in the nucleation mode, while the mass is dominated by the coarse and accumulation modes. It is perhaps unfortunate then that most current air quality metrics focus upon particle mass concentration (PM_{10} or $PM_{2.5}$), while scientific evidence suggests that many human health impacts are related to exposure to the fine-mode particles which can penetrate most deeply into the alveolar structure.

Measurements of the composition of urban particulate matter are less common than total mass concentration, but a number of composition measurements are now performed routinely at a small number of sites in various countries. Within the sphere of such semi-routine monitoring, the (mass) contributions from sulfate, nitrate, ammonium, chloride and other elemental components may be determined, alongside the elemental and organic carbon fractions. Figure 5, below, shows the composition of $PM_{2.5}$ and PM_{10} in Birmingham city centre. Components which are apparent include iron and calcium components from crustal

erosion, salt (sea-salt derived, possibly with some contribution from winter road salting), elemental carbon (essentially soot, from incomplete combustion; in these measurements, likely to arise almost entirely from diesel vehicles), organic compounds, ammonium bisulfate, ammonium nitrate and sodium nitrate. Of these, the ammonium, sulfate and nitrate components are likely to be largely secondary, arising from agricultural emissions, combustion of sulfur-containing fuels and the atmospheric processing of NO_x respectively (see above), while within the organic (carbon) fraction comprises elemental carbon (soot, likely primarily from diesel vehicles) and a substantial contribution from organic compounds.

Research-focused field measurements can disaggregate the organic fraction in more detail, and typically find that this is primarily secondary in origin (up to 64% in urban regions [30]), arising (in urban environments) largely from aromatic VOCs (e.g., benzene, toluene), which in turn are sourced primarily from vehicle emissions. As in the case of ozone, the substantial secondary component of urban particulate matter complicates air quality management and control. A component of particulate matter worth of further consideration is polycyclic aromatic hydrocarbons (PAH). Formed from two or more aromatic rings fused together, and produced from the incomplete combustion of coal, oil, petrol and wood, in particular



Urban Atmospheric Composition Processes. Figure 5

Chemical composition of particulate matter (PM_{10} and $PM_{2.5}$) from Birmingham City Centre. Data from Yin & Harrison [29]

domestic fuel usage, PAH are persistent bio-accumulative organic compounds; human exposure leads to a range of toxic and carcinogenic effects. One specific PAH, benzo-a-pyrene, is commonly used as a marker for the abundance of this family of compounds; the EU has established a target concentration of 1 ng m^{-3} for B[a]P by end of 2010, which may be compared with current (UK) levels of 0.16 ng m^{-3} (mean across all sites for 2006).

Trends in Urban Atmospheric Composition

This section concerns trends in the key urban atmospheric components over the forthcoming few decades. The trends observed reflect the integration of a number of independent effects: background levels of atmospheric components on a regional and global basis; changes in emissions of primary pollutants per unit activity (i.e., from improvements in vehicle technology and penetration of exhaust treatments such as catalytic converters through the vehicle fleet); and changes in the overall activity level, that is, increasing economic prosperity and population levels, particularly in many megacities in developing nations. A further external driver which may affect future urban composition is the interaction between atmospheric composition processes and global climate.

For the gaseous pollutants, the species of particular concern on an ongoing basis are nitrogen oxides (specifically NO_2 as this is the NO_x component against which legislation is commonly framed) and ozone. Of the other regulated pollutants (Table 2), sulfur dioxide levels are falling as the intensive sources (power and heavy industry) have emissions controlled at source and carbon monoxide levels have greatly reduced with the advent of catalytic converters.

Urban NO_x levels show a downward trend in most developed cities when viewed on a 2–3 decade timescale, driven by improved emission control technologies in the vehicle fleet; for example, across a range of UK urban sites, annual mean NO_x levels have fallen by 40% over the period 1993–2005, from 135 to $80 \text{ } \mu\text{g m}^{-3}$ [18]. Within this overall trend, there are some significant local variations, for example, data from London suggests that the fraction of NO_x emitted as NO_2 has risen in recent years, possibly due to the

introduction of diesel particulate filters, while traffic control measures such as the London Congestion Zone have altered the average fleet composition in central London to favor diesel vehicles with increased NO_x emissions, thought to be responsible for the 7% increase in NO_2 observed in central London over the period 2003–2008. More widely however, NO_x levels in most developed cities have fallen dramatically over the past 2–3 decades [18].

Globally, background ozone levels have risen over the past century – the earliest instrumental data thought to be reliable, the Montsouris measurements recorded on the outskirts of Paris from 1876, indicate average background ozone levels of $22 \text{ } \mu\text{g m}^{-3}$ over the period 1876–1910 [31]. The equivalent values for the past decade from background locations are much higher – for example, ca. $80 \text{ } \mu\text{g m}^{-3}$ for background air at Mace Head, Ireland (value for 2006 from Derwent et al. [32]); background levels have increased over the past 2–3 decades at up to $10 \text{ } \mu\text{g m}^{-3}$ per decade (5 ppb per decade) [33]. On top of this hemispheric trend, urban ozone levels respond to regional pollution events, and to the extent of local titration of O_3 to NO_2 through NO_x (strictly, NO) emissions. Emissions controls have reduced the prevalence and extent of ozone episodes over the past 1–2 decades, reducing this contributor to urban ozone; however reductions in NO_x emissions have offset some of this change, reducing the ozone urban decrement. These effects are clearly seen in the UK over the period 1990–2006 (e.g., AQEG [34]), where the maximum hourly mean values (associated with regional ozone pollution episodes) have decreased substantially due to European controls on emissions of VOCs and NO_x , while at intermediate percentiles, urban sites show a positive trend in ozone with time due to the reduced urban decrement. The lowest hourly mean ozone levels show no change as these essentially correspond to total titration of O_3 to NO, consequently ozone levels are zero across the time period considered.

Primary sources (vehicle exhaust) of PM have been substantially reduced in many developed nations, and PM levels (measured as PM_{10}) have fallen substantially over the past few decades. Recently however, the downward trend appears to have been arrested, at least for PM_{10} in Europe, and the sulfate and nitrate

components of this PM_{10} [35]. It is unclear if this is a temporary departure from the general trend or reflects an ongoing leveling off, possibly reflecting non-exhaust sources, changing meteorology, and the interaction between sulfate and nitrate via ammonia availability to form ammonium nitrate.

Future Directions

In many developing megacities, the outcome of the opposing effects of increasing population density and economic activity, and adoption of improved (i.e., lower emission) technologies for transport, industry and domestic heating/cooking will determine local atmospheric composition over the coming decade. Dramatic improvements have been shown to be possible in cities when suitably stringent emission controls are implemented, for example, during the Beijing Olympics in 2008 [14, 36]. A key issue for future urban atmospheric composition processes is the potentially changing climate within which the air chemistry takes place, and the global perspective of cities as sources of a range of pollutants into the wider atmosphere. A number of the processes outlined above are strongly temperature-dependent, in particular the stability of PAN and other organic nitrates which sequester NO_x , and the equilibria between the condensed and vapor phases for many of the secondary components of PM, in particular nitrate. Changed temperatures will also affect emission sources, including evaporative losses of fuels, and emissions of biogenic VOCs, many of which are strongly (and non-linearly) temperature-dependent, and has been suggested to affect the prevalence of the meteorological patterns which favor ozone pollution episodes, such as the western European heatwave of 2003 [37]. These latter effects may particularly influence the background atmospheric composition (ozone levels) upon which localized urban influences build. Urban heat island effects already drive some of these changes on a local scale. A potentially larger climate-related impact may however arise from the widespread adoption of non-fossil-fuel-based transportation systems, driven by carbon emission considerations. Should this occur, our currently low understanding of the non-exhaust contributions to urban air pollution will assume much greater importance.

Bibliography

Primary Literature

1. United Nations Department of Economic and Social Affairs/ Population Division (2008) World urbanization prospects: the 2007 revision. United Nations, New York
2. Fenger J (1999) Urban air quality. *Atmos Environ* 33:4877–4900
3. Mage D, Ozolins G, Peterson P, Webster A, Orthofer R, Vandeweerd V, Gwynne M (1996) Urban air pollution in megacities of the world. *Atmos Environ* 30:681–686
4. Brimblecombe P (1987) The big smoke. Methuen, London
5. Salmond JA, McKendry IG (2009) Influences of meteorology on air pollution concentrations and processes in urban areas. In: Hester RE, Harrison RM (eds) Air quality in urban environments, vol 28, Issues in environmental science and technology. Royal Society of Chemistry, London, pp 23–41
6. Thorpe AJ, Harrison RM, Boulter PG, McRae IS (2007) Estimation of particle resuspension source strength on a major London road. *Atmos Environ* 41:8007–8020
7. Molina LT, Molina MJ (2004) Improving air quality in megacities Mexico city case study. *Ann NY Acad Sci* 1023:142–158
8. Ding A, Wang T, Zhao M, Wang T, Li Z (2004) Simulation of sea-land breezes and a discussion of their implications on the transport of air pollution during a multi-day ozone episode in the Pearl river delta of China. *Atmos Environ* 39:6737–6750
9. Vardoulakis S, Fisher BEA, Pericleous K, Gonzalez-Flesca N (2003) Modelling air quality in street canyons: a review. *Atmos Environ* 37:155–182
10. Lee JD, Moller SJ, Read KA, Lewis AC, Mendes L, Carpenter LJ (2009) Year-round measurements of nitrogen oxides and ozone in the tropical north Atlantic marine boundary layer. *J Geophys Res* 114:D21302. doi:10.1029/2009JD011878
11. Read KA, Lee JD, Lewis AC, Moller SJ, Mendes L, Carpenter LJ (2009) Intra-annual cycles of NMVOC in the tropical marine boundary layer and their use for interpreting seasonal variability in CO. *J Geophys Res* 114:D21303. doi:10.1029/2009JD011879
12. United Nations Environment Programme (2007) Beijing 2008 Olympic games – an environmental review. Report of the United Nations Environment Programme. United Nations Environment Programme, Nairobi
13. Gros V, Sciare J, Yu T (2007) Air-quality measurements in megacities: focus on gaseous organic and particulate pollutants and comparison between two contrasted cities, Paris and Beijing. *C R Geoscience* 339:764–774
14. Wang T et al (2010) Air quality during the 2008 Beijing Olympics: secondary pollutants and regional impact. *Atmos Chem Phys* 10:7603–7615
15. Abam FI, Unachukwu GO (2009) Vehicular emissions and air quality standards in Nigeria. *Eur J Sci Res* 34:550–560
16. Raheem AMOA, Adekola FA, Obioh IO (2009) The seasonal variation of the concentrations of ozone, sulfur dioxide, and nitrogen oxides in two Nigerian cities. *Environ Model Assess* 14:497–509

17. World Health Organisation (2006) Air quality guidelines. Global update 2005. Particulate matter, ozone, nitrogen dioxide and sulfur dioxide. World Health Organisation, Copenhagen
18. Air Quality Expert Group (2007) Trends in primary nitrogen dioxide in the UK, Fourth report of the Air Quality Expert Group. DEFRA, London
19. Maynard RL (2009) Health effects of urban pollution. In: Hester RE, Harrison RM (eds) Air quality in urban environments, vol 28, Issues in environmental science and technology. Royal Society of Chemistry, London, pp 108–128
20. Jacobson MZ (2007) Effects of ethanol (E85) versus gasoline vehicles on cancer and mortality in the United States. *Environ Sci Technol* 41:4150–4157
21. Andrae MO, Merlet P (2001) Emission of trace gases and aerosols from biomass burning. *Glob Biogeochem Cycles* 15:955–966
22. Leighton PA (1961) Photochemistry of air pollution. Academic, New York
23. Bloss WJ (2009) Atmospheric chemical processes of importance in cities. In: Hester RE, Harrison RM (eds) Air quality in urban environments, vol 28, Issues in environmental science and technology. Royal Society of Chemistry, London, pp 42–64
24. Wayne RP (2001) Chemistry of atmospheres. Oxford University Press, Oxford
25. Jenkin ME, Davies TJ, Stedman JR (2002) The origin and day-of-week dependence of photochemical ozone episodes in the UK. *Atmos Environ* 36:999–1012
26. Emmerson KM, Carslaw N, Carslaw DC, Lee JD, McFiggans G, Bloss WJ, Gravestock T, Heard DE, Hopkins J, Ingham T, Pilling MJ, Smith SC, Jacob M, Monks PS (2007) Free radical modelling studies during the UK TORCH campaign in summer 2003. *Atmos Chem Phys* 7:167–181
27. Wood EC, Herndon SC, Onasch TB, Kroll JH, Canagaratna MR, Kolb CE, Worsnop DR, Neuman JA, Seila R, Zavala M, Knighton WB (2009) A case study of ozone production, nitrogen oxides, and the radical budget in Mexico city. *Atmos Chem Phys* 9:2499–2517
28. Derwent RG, Jenkin ME, Saunders SM (1996) Photochemical ozone creation potentials for a large number of reactive hydrocarbons under European conditions. *Atmos Environ* 30:181–199
29. Yin J, Harrison RM (2007) Pragmatic mass closure study for PM_{1.0}, PM_{2.5} and PM₁₀ at roadside, urban background and rural sites. *Atmos Environ* 42:980–988
30. Zhang Q et al (2007) Ubiquity and dominance of oxygenated species in organic aerosols in anthropogenically-influenced northern hemisphere midlatitudes. *Geophys Res Lett* 34: L13801. doi:10.1029/2007GL029979
31. Volz A, Kley D (1988) Evaluation of the montsouris series of ozone measurements made in the nineteenth century. *Nature* 332:240–242
32. Derwent RG, Simmonds PG, Manning AJ, Spain TG (2007) Trends over a twenty year period from 1987–2006 in surface ozone at the atmospheric research station at Mace Head, Ireland. *Atmos Environ* 41:9091–9098
33. Raes F, Hjorth J (2006) Answers to the Urbino questions. ACCENT Secretariat, Università di Urbino, Italy. ISBN 92-79-02413-2
34. Air Quality Expert Group (2009) Ozone in the United Kingdom. Fifth report of the Air Quality Expert Group, DEFRA, London
35. Harrison RM, Stedman J, Derwent D (2008) New directions: why are PM₁₀ concentrations in Europe not falling? *Atmos Environ* 42:603–606
36. Fang M, Chan CK, Yao X (2009) Managing air quality in a rapidly developing nation: China. *Atmos Environ* 43:79–86
37. Lee JD et al (2006) Ozone photochemistry and elevated isoprene during the UK heatwave of August 2003. *Atmos Environ* 40:7598–7613

Books and Reviews

- Ayres J, Maynard RM, Richards R (eds) (2006) Air pollution and health, London: Imperial College Press reports of the air quality expert group. DEFRA, UK
- Brimblecombe P (1987) The big smoke. Methuen, London
- Brimblecombe P, Maynard RL (2001) The urban atmosphere and its effects, vol 1, Air pollution reviews. Imperial College Press, London
- Fenger J (1999) Urban air quality. *Atmos Environ* 33:4877–4900
- Harrison RM, Hester RE (eds) (2009) Air quality in urban environments, vol 28, Issues in environmental science and technology. Royal Society of Chemistry, Cambridge
- Holloway AM, Wayne RP (2010) Atmospheric chemistry. RSC, Cambridge
- Singh HB, Brimblecombe P, Veenstra F (2009) Atmospheric environment special issue fifty years of endeavour and papers therein. *Atmos Environ* 43:1
- Tiwary A, Colls J (2010) Air pollution: measurement, modelling and mitigation, vol 3. Routledge, London
- WHO/UNEP (1992) Urban air pollution in megacities of the World. Blackwell, Oxford
- WHO (2006) Air quality guidelines. Global update 2005. Particulate matter, ozone, nitrogen dioxide and sulfur dioxide. World Health Organisation, Copenhagen

Urban Ecology

S. T. A. PICKETT¹, M. L. CADENASSO²

¹Cary Institute of Ecosystem Studies, Millbrook, NY, USA

²Department of Plant Sciences, University of California, Davis, CA, USA

Article Outline

Glossary

Definition of the Subject

Introduction
 Urban Ecology Principles
 Future Directions
 Bibliography

Glossary

City A dense, demographically and economically heterogeneous settlement containing businesses, residences, intensive transportation infrastructure, multistory buildings, and warehousing or manufacturing. Economies of cities focus on services, processing commodities, manufacturing, or finance rather than on agricultural or resource management. Cities are often characterized by a population representing diverse social groups and economic classes in relatively close spatial proximity. As a caution, it should be noted that the term “city” is sometimes used to refer to an entire urban (definition 1) area, and sometimes used in a narrow sense to contrast with suburb (definition 2).

CSE City-suburban-exurban system. A comprehensive term referring to all components of a complex urban system, metropolitan area, or other cluster of urban areas.

Ecology The science of the relationship of organisms to each other and the physical environment, and the transformation of resources mediated by those interactions. Human ecology includes the social and institutional structures established by people as components of the system studied.

Ecosystem A system comprising the organisms, the physical environment, and the interactions among them within a specified volume of the Earth. Ecosystems may be of any size, depending on the research questions of interest, and are open to material and energetic flows with adjacent systems. Although this definition encompasses human ecosystems, for completeness it is considered that human ecosystems include biological components, the physical environment of air, water, soil, energy, the social and human institutions, and the built environment.

Suburb This term has two meanings, depending on whether an Old World or a New World context is intended. In the New World, a suburb is a component of a broadly recognized urban area, a primarily

residential land cover in which single or multiple household dwellings are interspersed with the open spaces of lawns and generous street landscaping. New World suburbs are often a locus of wealth and power. Modest commercial nodes, mostly as service rather than manufacturing or warehousing, may be included in such suburbs. In contrast, Old World suburbs may be less green and more densely built than New World suburbs, and in many cases may host concentrations of lower income, less empowered persons. In the expanding cities of the global south, shanty towns and informal settlements may constitute much of the suburban realm.

Urban Definition 1 refers to all components of densely settled and built up areas, and contrasts with rural, agricultural, or wild lands. An alternative terminology for this inclusive definition is city-suburban-exurban (CSE) system. Definition 2 refers to dense commercial, industrial, and residential lands in contrast to suburbs and exurbs.

Definition of the Subject

Within the science of ecology, urban ecology is defined as the study of structure, dynamics, and processes in urban ecological systems. Urban ecology is the study of the relationships of human and nonhuman organisms in urban areas, the interactions of these organisms with the native and built physical environment, and the effects of these relationships on the fluxes of energy, materials, and information within individual urban systems and between urban and nonurban systems. Urban ecology applies the methods and concepts of the biological science of ecology to urban areas, but requires and integrates with the concerns, concepts, and approaches of social sciences to produce a hybrid discipline. Urban ecological systems include individual organisms, populations, communities, and landscapes, as well as buildings and infrastructure. Urban ecology further recognizes specific urban ecosystems as a part of the global biogeochemical, economic, and human demographic system.

Importance of Urban Ecology

Urban ecology is important because it brings the insights and knowledge from contemporary biological ecology to bear on urban areas [1]. It replaces the

earlier and superseded versions of ecological science that had been used by social scientists, geographers, and urban planners in justifying and predicting urban dynamics in the second half of the twentieth century. Urban ecology as a branch of contemporary ecological science now emphasizes spatial heterogeneity, feedbacks between natural and human system components, probabilistic system change, and the integration between human perceptions and environmental processes.

Urban ecology is also important because urban habitats are increasing worldwide. The United Nations estimates that more than 50% of the global population now resides in urban areas, as defined by the various member nations. In addition, the next three billion people to be added to the world population are expected to live in urban areas. Hence, urban systems are becoming the predominant habitat of humanity, and are an increasingly widespread land cover type worldwide. In the USA, constructed surfaces now cover an area equivalent to that of the state of Ohio [2].

If the disciplines and practices of urban planning and design, ecological restoration, and ecosystem management are to draw more effectively upon ecological knowledge and data, then the science of urban ecology will become an increasingly key resource for these pursuits.

Brief History

Urban ecology has emerged as a subdiscipline of biological ecology only in the last 30 years [3]. It began as an ecological science in the study of the species and biotic communities of conspicuously green patches in cities and metropolises. Parks, vacant lots, disturbed, and derelict lands were the first focal areas of the discipline [4]. More recently, ecologists began to examine areas actively inhabited and managed by people, including lawns and streetscapes [5]. Another contrasting tradition in urban ecology focuses on the coarser scale, to quantify energy and material budgets of cities. This focus, sometimes called urban metabolism, deals with the energy, matter, and information that flow through and are transformed by cities. In all cases, how the biological components and fluxes affect the well-being of people in the city is a concern. However, the contemporary approach to urban ecology

differs from the past traditions. First, all areas in the city are now subject to ecological analysis, not just the conspicuous green areas. Second, even in the budgetary approach, the internal spatial configuration of different components of the urban area is recognized as potentially influencing the fluxes and transformations within the larger metropolis. Finally, the fully hybrid nature of the systems is acknowledged, so that cities are seen as neither fully human nor fully natural entities. Rather, they are inextricably both human constructions and biophysical features [6, 7]. Urban ecology was once a study of green spaces in the city. Now it is the study of the ecology of the entire urban area, including biological, built, social, and physical components.

Other scholarly disciplines beyond biology have used the term “urban ecology.” Principal among these has been sociology. This use originated in the 1920s at the University of Chicago under the leadership of Robert Park and Ernest Burgess, who brought concepts of community, competition, and succession that were then current in biological ecology into their new discipline of sociology. Human ecology, which has roots in geography, anthropology, and other social sciences, is closely related to urban ecology when the study subject is urban populations and their interactions. However, other disciplines tend to neglect the physical and biological components of the environment when they address urban ecology.

Introduction

Urban ecology has been used by several disciplines which have different foci and concerns. These contrast or complement the conceptions of urban ecology as a biological science, which is the approach emphasized here.

History

Urban ecology has a long history. The first flowering of urban ecology was a sociological phase established by Park and Burgess at the University of Chicago in the 1920s. Although this was a sociological pursuit, it was centrally informed by analogies from the biological science of ecology, for which the University of Chicago was one of the founding schools. Park, Burgess, and their students explained the unprecedented growth and social change in Chicago in terms of invasion of new human communities, competition among

communities, and spatial isolation between different communities and functions in the city. These scholars were disturbed by the doubling of the population of Chicago at the time, and the role of new migrants from the American South or from eastern and southern Europe. The racial, ethnic, and class novelty in the city begged explanation and incited the Chicagoans to seek explanatory and predictive models to serve in the face of such unprecedented changes. This approach to urban ecology was informed by a tacit antiurbanism, as the Chicago sociologists held village and agricultural communities as the paragons of human societies. One of the central tenets of the Chicago school was that cities had a life cycle, analogous to the expected, but incorrect, prediction that ecological communities had predictable life cycles starting from invasion, extending through competition and sorting, and ending in a mature state. This phase of urban ecology ended when social science critics prompted a move toward more individual behavioral explanations of urban change, as opposed to community-based models. A similar, but independent shift occurred in mainstream ecology at about the same time. Even though the academic community moved beyond the deterministic, life-cycle approach to cities, urban policy in the USA continued to assume life-cycle patterns through the 1960s, basing urban conservation and urban renewal policies on this flawed assumption.

Oddly, during the early twentieth century, while their major ideas were informing the birth of sociology and being widely applied in urban systems, most biological ecologists heartily ignored cities and urban systems. European and Japanese ecologists began to explore ecology in urban contexts after World War II. The manifest destruction in the cities in which they lived invited their interest as biologists. What would be the patterns and mechanisms of plant establishment in derelict sites? How would the newly established biotic communities change over time? What benefit might they provide the cities in which they occurred? The questions of the immediate postwar researchers in Europe and Japan were standard ecological questions, but asked in a novel location. This tradition became linked with urban planning in Europe and has remained active in that form [8].

The second wave of urban ecology rose in the 1970s in the USA. Associated with the birth of

environmentalism and its concern with the Earth's exponential human population growth, the urban ecology of this era tended to assume that humans were a negative influence on ecosystems, and urban areas provided an extreme case of the human impact that was beginning to worry scientists and the public. A key document from this era is the volume by Stearns and Montag [9]. In it, the problems of urban areas are outlined, and the nature of potential ecologically informed solutions is suggested. However, the ecology of the time was rather coarse-scaled, and assumed equilibrium tendencies of systems, rather than recognizing fine-scale heterogeneity as a causal feature of systems [10]. Furthermore, although failure of the old ecological ideas that had informed the Chicago School was evident, no clear replacement had emerged. Urban ecology in this era concentrated on investigations of conspicuously green patches in the city. Hence, this approach can be characterized as ecology in the city [3]. Parks, cemeteries, gardens, and abandoned lots exemplify this literature.

Another feature of this second wave of urban ecology was a budgetary, systems approach. Epitomized by work in Hong Kong [11], this approach to urban ecology addressed energy and material budgets of cities, and detailed the human costs of pollution and crowding. This approach is characterized as a budgetary feature of ecology of the city. It shares with the early Chicago School an assumption of the importance of urban "pathologies" in the human population. Industrial ecology and urban metabolism are branches from this tradition. Both of these schools of thought analyze the material and energetic inputs, efficiencies, and outputs of urban systems and their components. Life-cycle analysis of materials is a strategy that aims to reduce the use of resources and the generation of wastes associated with contemporary material use. This era of urban ecology did not persist in the USA as a comprehensive field.

A new wave of urban ecology is currently on the rise. It is characterized by several features that differentiate it from prior instances of urban ecology, and make it more comprehensive than earlier approaches. First, it attempts to unify social and biological knowledge, concerns, and approaches [12]. Second, it acknowledges and exploits spatial heterogeneity and fine-scale dynamics as a feature and cause of urban change.

Third, it seeks to understand the controls of biogeochemical processes throughout urban systems, including retention, fluxes, and leakage of limiting nutrients and pollutants. Contemporary urban ecology brings the three previously separate goals together for the first time.

Will this current interest in urban ecology wane, as did the previous ones in the USA? One difference between the current manifestation of urban ecology and the previous ones is institutional support. The pioneers of urban ecology in Europe, Japan, and the USA did not have long-lasting research support. As a result, their pioneering efforts were sometimes short-lived. Now there are two urban Long-Term Ecological Research (LTER) sites in the USA, and International Long-Term Ecological Research programs and Zones Ateliers are including urban areas among their rosters. Already the US LTER urban sites are 13 years old. Such longevity promotes interdisciplinary collaboration, continued use of research areas, developing ongoing relationships with communities and decision-makers, and accumulation of lengthy data runs which can expose causal links and the role of pulse events [13]. Acknowledging that urban areas both contribute to and are vulnerable to global changes [13] will tend to keep them in focus in ecological science.

Examples

Urban ecology is such a diverse science that examples are required to give a sense of its breadth.

Patterns of diversity and abundance associated with urbanization are complex and competing explanations exist. Tests of island biogeography theory in urban areas find that species–area relationships are preserved in urban patches [14]. However, in some studies, patch size influenced species composition rather than species richness as a result of organisms at higher trophic levels being preferentially lost from smaller patches [15]. Attempts to directly quantify the extinction and colonization processes that island biogeography relies on have shown that immigration and extinction characterize different kinds of patches [16, 17]. The species composition in a patch is the result of species colonizing the novel habitats formed by urbanization along with those remaining after local extinctions due to

isolation or habitat alteration. One prediction of the view of complex causes of urban biodiversity is that urban habitats are not always less diverse than rural patches. Rather, diversity depends on the sum of extinction and colonization rates, which differ regionally and taxonomically. At moderate levels of urbanization, species richness may actually be higher than in nearby wild lands.

A second example is the disconnection between riparian zones of urban streams and the water table [18]. This disconnection limits the capacity of urban riparian zones to convert nitrate, a pollutant in groundwater derived from fertilizer and sewage, to gaseous forms that do not pollute streams. Research in agricultural landscapes has suggested that riparian restoration, inserting woody and grass vegetation between crops and stream banks, is an effective strategy to mitigate nitrate pollution in streams. When the capacity of urban riparian zones to accomplish such mitigation was examined in Baltimore, MD, USA, it was discovered that riparian zones had become disconnected from the groundwater sources that control their ability to convert nitrate to nitrogen gas. With reduced infiltration of stormwater into the ground due to impervious surfaces, and with high incision leaving stranded droughty floodplains in cities, urban riparian zones no longer support the anaerobic conditions and high organic matter required to fuel denitrifying bacteria. Hence, the expected denitrification in urban riparian zones may not always occur [19]. This example demonstrates that knowledge obtained in nonurban environments may not apply to urban situations.

Guide to the Article

It is now clear that urban areas express different combinations of ecological processes than do nonurban areas and that urban areas are an increasingly important component of the global biosphere. It is now possible to examine key principles that are emerging from the new ecological focus on urban research. Future directions for research and management are addressed at the conclusion of the article.

Urban Ecology Principles

Thirteen principles characterize the contemporary science of urban ecology. Of course, such principles are

likely to be improved or replaced with advances in this rapidly growing scientific field. The principles can be divided into four groups: (1) the human ecosystem, (2) urban form, (3) urban function, and (4) methodology.

Principles Concerning Human Ecosystems

Much of the history of urban research has proceeded as though city and ecology were different and mutually exclusive. At the same time that the social scientists of the Chicago School were applying bioecological concepts of the day to their work in the cities, another member of the University of Chicago faculty was establishing the concept of plant succession. A member of the Botany Department, Henry Chandler Cowles, was working on the seemingly pristine plant communities of the Indiana Dunes, distant from Chicago's immigrant-driven hurley burley. There was conspicuously no empirical or theoretical collaboration between the pioneering ecologists and social scientists at the University of Chicago in the early twentieth century [20].

The tradition of treating city and nature as opposites was challenged by adventuresome scholars and practitioners in the closing decades of the twentieth century [21]. This has led to a new conception of cities or urban areas as hybrid socio-bioecological systems. In other words, urban places may be considered to be human ecosystems. As such, they incorporate not only the traditionally recognized biotic and physical components of ecosystems, but also the social structures and built components so conspicuous in cities and towns [7].

The new conception of human ecosystems prompts researchers, planners, and managers in urban systems to study and exploit the reciprocal feedbacks between the social components in all their demographic, institutional, behavioral, and economic complexity, and bioecological processes, whether conspicuous or not. For example, the mitigation of urban heat extremes by trees is well known [22]. Put simply, there is an ecosystem feedback between vegetation and human comfort or risk of heat stress. However, because of the multifaceted nature of the human ecosystem, the embedding of human values and culture, and the linkage of vegetation cover with water demand, the causal link

between the desire to mitigate heat stress and the willingness to plant trees is complex. Included are citizen concerns over private property access, commitments to tree maintenance and litter removal on their property, fear of crime associated with vegetated hiding places, risk of treefall, root fouling of infrastructure, and aesthetic judgments [23]. Other examples of such processes are associated with later principles.

Two principles emerge from these considerations about human-natural system coupling in urban contexts:

Principle 1: Cities and urban areas are human ecosystems in which socioeconomic and bioecological processes feed back to one another.

Principle 2: Human values and perceptions are a key link mediating the feedbacks between social and bioecological components of human ecosystems.

Principles Concerning Urban Form

One of the most remarkable features of urban areas is their spatial form. Architects and urban planners have traditionally used figure-ground representations of urban areas, which emphasize the built component of the human ecosystem. Pioneering efforts by landscape architects have shown the need and power of going beyond such classic representations to include both the hidden and the conspicuous bioecological features of urban areas.

Urban form has changed dramatically through time. Early walled cities were physically and socially distinct from the surroundings and often were centered on a cosmologically significant, ceremonial core. The industrial city spread beyond the location of former city walls and established porosity for the purposes of exchange of goods and the accumulation of immigrants needed as labor. After World War II, the industrial city was further disaggregated into a mixture of new suburbs, dispersed commercial districts, and older core cities that were reduced in density. This is the city as part of a network in a dispersed megalopolis [24].

This simple, linear typology does not necessarily reflect the temporal trends in all parts of the world. Indeed in some cities of the global south, no industrial period existed before they began the sprawling growth driven by migration of persons from the provinces seeking opportunity and a better life in the modern city of consumption. Shanty towns outpacing the urban infrastructure are one feature of such cities.

Hence, urban form is evolving, which leads to the following principle of urban form:

Principle 3: Urban form is a dynamic phenomenon and exhibits contrasts through time and across regions that express different cultural and economic contexts of urbanization.

Why is urban form important ecologically? The establishment of street grids and major road patterns alters surface drainage [25] and presents barriers and corridors for the movement of native and introduced plants and animals [26]. The adding of new vegetation and the obliteration of the forest, savanna, desert, wetland, or grassland that had previously occupied the urban site follows the dictates and needs of the new urban form. In addition, the infrastructures for supplying clean water and for dealing with fouled water are large alterations to watershed structure and regional water balances. The placement of main sanitary sewers and main storm drains in stream valleys and the shortcutting of ground and surface water flows are components of the universal urban alteration of hydrology. Furthermore, the shapes of heterogeneity introduced by urban form are often rectilinear and of different scale than the former regional and local patterns of environmental gradients and patches. In addition to spatial heterogeneity of the biophysical components of urban systems, the social features of cities, towns, and suburbs are notably patchy [27]. Urbanists have long acknowledged the fine-scale change – often from block to block – in economic activity, wealth, social group, architecture, and land use that characterize cities. Although traditional postwar suburban development typically occupies rather large tracts which they homogenize, the abrupt shifts in urban structure persist. This patchiness is reinforced by the near universal employment of zoning in large urban areas of wealthy countries. All of these heterogeneous structural alterations can have significant effects on urban and adjacent natural ecosystem function, as will be detailed in a later section “[Principles Concerning Urban Function](#)”. These insights about the nature of urban form at scales as small as residential parcels or a stroll to the corner of the block can be summarized in a principle:

Principle 4: Urban form is heterogeneous on many scales, and fine-scale heterogeneity is especially notable in cities and older suburbs.

The heterogeneous form of contemporary urban systems extends into the surroundings [28]. The suburban fringe in rapidly growing cities may abut farms, desert, forest, or whatever landscape constitutes the predominant rural or wild land cover. Indeed, the adjacencies are often convoluted and complex, leading to interdigitation of urban elements with the wilder or more rural landscape. In some wealthy urban regions, individual homes are embedded in what is otherwise a wild matrix [29]. An example is the building of suburban houses in chaparral shrublands in Mediterranean climates, such as Southern California. A similar case is the construction of houses in the forest fringe of the Pacific Northwest metropolises, or in forested foothills of the exurban lands of the Southern Appalachians in the USA. The insertion of households whose financial equity, lifestyle identity, social connections, and environmental attitudes have been defined by urban life rather than a life of logging, farming, fishing, or other natural resource management is perhaps a more significant interdigitation than the mere presence of their homes. Similar juxtapositions exist for resorts that serve as remote summer or winter destinations for urban dwellers. The interdigitation of urban with wild or less-intensively managed lands results not only from the invasion of wild land by new housing, commercial, or transportation corridors, but also by the afforestation of foothill developments in arid or semiarid climates in which the lower slopes would have been savanna or grassland. The well-to-do suburbs of Oakland, CA, are an example, which now merge to some extent with the ridgetop forests of the Coast Ranges, and where fire is now a real risk. The fact that urban land cover is growing more rapidly than urban population in oil-subsidized economies [24] suggests that such interdigitation will be the source of increasing conflicts and changing exposure to natural disturbance regimes [29]. Recognizing the ongoing interaction of urban and wild lands suggests this principle:

Principle 5: Urban land covers and uses extend into and interdigitate with rural or wild land covers and uses.

Urban form and hence its interaction with less-urbanized landscapes is the result of a complex of causes. One cause of urban form is conformity to regional plans. This cause is rare in many countries, such as the USA, but more common in Europe.

Successful constraint on urban sprawl is exemplified in the USA by the metropolitan green line in Portland, OR, or the Urban-Rural Demarcation Line of Baltimore County, MD [30]. In Baltimore County, more than 90% of residents live within the URDL. The line has been in force since the 1970s and is controlled through zoning, the restriction of sewer infrastructure, and the focusing on urban growth in the county to specific corridors. However, the rarity of such regional controls is compensated for by the commonness of the role of developer self-interest in determining urban form. Indeed, urban planners often lament the predominance of the real estate market as a driver of urban form. On the social roster of causes of urban form are those that are described as “push and pull” factors. Persons and households choose to live at specific places in the metropolis based on such pull factors as attraction to open space, commodious housing, affordability, good schools, and access to desirable commercial establishments. Push factors include high taxes, small and old housing, lack of personally controlled open space, perceived risk of crime, underperforming schools, and the like. Firms also experience push and pull factors. In the case of businesses, a customer base, access to efficient delivery routes, free parking, access to a trained work force, and similar concerns are pull factors. The push factors for businesses mirror those experienced by households.

Many of the factors that control urban form are accidental or unintentional. Pollution, traffic congestion, and the concentration of heat in built areas are examples of unintentional disamenities. Urban flooding is an unintended result of successfully draining stormwater from upstream. Drawing down of groundwater is a similar result of successful stormwater drainage. In neighborhoods built on filled wetlands in Boston, such drying of the soil is threatening the wooden piles supporting the foundations of some buildings. These examples suggest attending to both the intended and unintended consequences of urban development and environmental modification. The principle that emerges is the following:

Principle 6: Urban form reflects planning, incidental, and indirect effects of social and environmental decisions.

Because ecology as a science has been absent from cities for so long the need for data is great. However, ecologists cannot rely on the tried and true experimental

approach which they employ with such enthusiasm and success in more wild and rural lands. In urban areas, multiple ownerships, the involvement of individuals and communities in land-use decisions even on private land, and the ethical constraints about manipulations that affect people’s well-being or pleasure all limit the applicability of controlled, replicated experiments.

However, while ethics and property regimes close one door to experimentation, the processes of urban design, development, and neighborhood revitalization open others [31]. Urban design projects, whether they be landscape architecture projects in the public realm, private development of a subdivision, or installation of environmental retrofits in older residential or commercial areas, can themselves be treated as experiments. Two examples show the power of this approach. One, the classic case of Jordan Cove, CT, compared an alternative suburban development to a traditional layout of homes. The goal was to test the potential to mitigate stormwater quality and amount with minimal effect on real estate development. A second example is the proposed subdivision of a forested tract in the Hudson Highlands region of New York State. The goal in this case was to reduce the impact on a salamander population that is of conservation concern in the State. The landscape architect employed ecological principles to suggest altering the traditional suburban design to reorient roads, preserve small vernal ponds upon which the salamanders depend for breeding, and install rain gardens as a component of the yard of each house to maintain the hydrologic regime of the complex network of ponds and overland drainage. The development is planned to contain drift fences and pitfall traps as infrastructure to continually monitor the salamander population. Hence the development itself can act as an experiment. To work with designs as experiments, partnerships with urban designers and developers will be required. These examples suggest a principle to test in further application:

Principle 7: Urban designs and development projects at various scales can be treated as experiments to expose the ecological effects of different design and management strategies.

Principles Concerning Urban Function

The tradition in both the social and biophysical sciences to see urban areas as distinct and opposed to

nature has resulted in a bias against finding natural processes in City-Suburban-Exurban systems. Both the pioneering research in landscape architecture and urban ecology [5] as well as the more recently established projects [13] have, however, confirmed and extended the understanding of ecological processes as parts of cities and their more extensive urban mosaics. Several examples show the power of bioecological phenomena in urban contexts.

In Philadelphia and Boston, buried floodplains, thought to have been engineered in a reliable stormwater management structure or filled to provide substrate for building, continued to manifest higher levels of soil water and were associated with property damage and abandonment [32]. In Philadelphia, the collapse of the sewer that had replaced Mill Creek was a catastrophic outcome of ignoring the function of the urban landscape. Landscape designs envisioned with the participation of neighborhood residents account for hazard and vulnerability, and provide an opportunity to convert a disturbance prone site to an open space amenity. Such sites can contribute constructively to regional stormwater management, recreation, and education.

A second example shows the capacity of open space to perform biogeochemical ecosystem functions in the urban matrix. Nitrate pollution in streams can be used as an index of environmental quality to assess ecosystem function. Researchers who started the Baltimore Ecosystem Study expected suburban lands to be detrimental to the environmental quality of the urban mosaic. This hypothesis was suggested by the large amounts of nitrogen fertilizer and water applied to many American lawns. When small watersheds that drain areas having different amounts of green space in their landscapes were studied however, it was discovered that the suburban subwatersheds exhibited relatively high nitrate retention [18, 33].

A third example of the role of ecological processes is shown in the net carbon flux between surface and atmosphere. In a location near the boundary between Baltimore City and County, an atmospheric flux tower assesses the upward versus downward movement of moisture, temperature, and carbon dioxide, among other factors. Although the urban area is a net generator of carbon dioxide to the atmosphere due to the use of fossil fuels for heating and transportation, there is

lower net flux on weekends and when the wind blows across areas of higher vegetation cover compared to winds from areas of greater built cover.

Examples such as these suggest a principle:

Principle 8: Urban areas contain remnant or newly emerging vegetated and stream patches that exhibit bioecological functions.

One of the fluxes that exercises the interest of urban planners and managers worldwide is water. This concern extends well beyond those cities that are located in arid or semiarid environments. Indeed, some cities, such as Phoenix, AZ, and Los Angeles, CA, are so well connected to distant water supplies via diversions from the Colorado River that they may seem impervious to drought. Ironically, although arid-land cities do show concern for water, many cities from moist climates are coming to acknowledge the sensitivity of their reservoir systems to periodic drought [34]. New York City, with its system of impoundments more than 200 km to the north, or Boston with its reservoir in rural central Massachusetts must now plan for the impacts of periodic drought and of the drying that may come with climate change. Even now, development in some counties in eastern Maryland faces limits due to shortages of groundwater, which are the sources for these suburban-style developments.

Such concerns with water supply are not limited to the USA. Urbanization on the Indian subcontinent is sensitive to projected reduced snowpack in the Himalaya Mountains, in which its rivers originate. Water withdrawal for agriculture and forestry in highlands of South Africa stands to alter water availability for which lowland settlement and wildlife compete. Similar diversions along the tributaries of the Nile are of concern for international relationships in East Africa.

Water supply is not the only aspect of water flux that is of concern. All cities, no matter whether they are in moist or dry climates, must deal with periods of high rainfall. Even desert cities are subject to occasional local thunderstorms, or to floods from the foothills and mountains upstream. Hence, stormwater management in spates is as of as much concern in Mediterranean Santiago, Chile, as it is in temperate Seattle, Washington. The traditional mode of stormwater management is to collect it and pass it off downstream as rapidly as possible. While this solves

local problems, it creates problems downstream. Stream bed erosion, scouring out the habitats of aquatic organisms and the spawning grounds of anadromous fish, and increasing the temperature of downstream reaches affected by runoff from hot paved surfaces are some of the consequences to the engineering approach to stormwater management. Many urban management plans now reference water supply and stormwater management as key concerns.

A principle emerges from these sorts of events and situations:

Principle 9: The flux of water, including both clean water supply and stormwater management, is of concern to urbanizing areas worldwide, and connects them explicitly to larger regions.

Ecologists and other biologists have recognized the presence of wildlife and plants in cities for a long time. This tradition is especially strong in Europe, but also has early roots in Japan and the USA. Interest has focused both on organisms that represent the native biota and those that represent threats to either native biota or to human health.

Native biological diversity, or biodiversity, is usually reduced in urban areas, with native species declining in richness compared to species introduced from elsewhere [35]. An example of altered species distribution that may have an impact on the future composition of urban forests is the greater number of vine species, especially introduced species, in the treefall gaps found in urban compared to rural forests. Such vines can impede the regeneration of trees in canopy gaps. Still, the total number of species, including both native and exotic species, can be larger in urban areas than in nearby rural areas in the temperate zones. However, in some central German cities, the richness of even native species is greater in urban areas than in the adjacent countryside. Urban wildlife is subsidized by purposeful feeding, by the incidental increase in resources found in garbage, or for fruit-feeding birds, by the large abundance of fleshy-fruited horticultural species planted in settlements [36].

Urban fauna and flora can respond to the novel conditions of urban ecosystems. For example, bird populations in San Diego exhibit genetically different plumage from rural counterparts [37]. Although that difference is not necessarily adaptive, other changes are functional. Urban bird populations have been

demonstrated to raise pitch and volume of their songs, presumably selected by the higher noise levels in cities than in the countryside [38]. Plants are well known to genetically adapt to the heavy metals found in brownfields or associated with urban construction.

An important question is the contribution of exotic and native species to ecosystem services in urban areas. Biophilia, the affinity that humans express for natural settings and living things or the salubrious effects of plants and wild animals in the surroundings, are benefits of wildlife and plants in cities. Other services are performed by the biota of cities, including moderation of climate through shading and transpiration, stabilization of soil surfaces and stream banks, and absorption of pollution. These services can be performed by both introduced and native species. Indeed, some introduced species, such as the exotic Norway maple, were imported due to their tolerance of urban stresses. Ironically, this species is one that has escaped into the wild, and due to the deep and seasonally long-lasting shade it casts, poses a threat to the herbaceous species of the forest understory and to regeneration of other canopy tree species [39].

Because of the damage that exotic species can do to native biotic communities and the functional changes they can cause in ecosystems, they are often targeted for removal. Although the introduction, either purposeful or accidental, of new exotic species should be avoided, in urban areas the removal of exotics as a blanket policy may be problematic. The contribution of the species to ecosystem services in urban areas, the ability of native species capable of performing that function to tolerate urban conditions, or the potential to breed native species for these capacities are all issues that must be considered in planning the management of exotic species. Some exotics can contribute to ecological functions in city-suburban-exurban habitats.

A principle that emerges from knowledge about urban biota is this:

Principle 10: Urban biodiversity has multiple components, and the contributions of each kind of species to ecosystem services in cities must be evaluated.

One important aspect of urban function is how equitably the ecological benefits and risks in urban systems are distributed across the human population. This concern, labeled environmental justice, first emerged in communities of color or economically

disadvantaged communities that were adjacent to contaminated sites, or sites that produce hazardous pollution. The frequent locations of such communities in topographic positions subject to flooding or landslides are other cases of environmental injustice. An example is the disproportionate damage suffered by low-lying neighborhoods such as the lower 9th Ward of New Orleans during 2005 in the wake of Hurricane Katrina, or the location of informal settlements with their flimsy construction in some South American cities on erodible slopes. Poor and disempowered populations may be actively excluded from desirable sites, as was the case with segregation practices historically in some US cities, or they may be indirectly priced out of locating to sites on higher ground, more distant from polluting businesses, or from facilities that process waste. The patterns and mechanisms of environmental injustice suggest a principle:

Principle 11: Bioecological processes are differentially distributed across the metropolis and the limitation of services and excess of hazards is often associated with the location of populations that are poor, discriminated against, or otherwise disempowered.

Methodological Principles

Two methodological principles of urban ecology complete the survey of guiding concepts for contemporary urban ecological science. These principles draw on conceptual and empirical principles outlined above and can help shape future urban research.

The first methodological principle deals with setting the boundary and scope of the study system. This step is key to quantification and comparison within and across urban systems. However, there is no a priori, universally correct scientific specification of an urban system [40]. Using legally set municipal or town borders may be convenient for policy purposes, but they may not be appropriate limits for scientific research [28]. This is because a large urban system may be made up of many such legally distinct jurisdictions. In addition, materials, energy, and information that are important in structuring an urban system are likely to move across political boundaries. For example, while it is possible and useful to measure species richness in a city, the collection of species so identified may depend on and interact with the larger, regional species

pool [14]. A second example compares boundaries chosen for watershed-based studies with those chosen for a policy analysis. The 17,150 ha Gwynns Falls Watershed, which extends from its headwaters in Baltimore County to its confluence with Baltimore Harbor near downtown in the City of Baltimore, provides a sampling transect. Measurements of stream water quality and flow integrate across the two jurisdictions [18]. In contrast, measurements intended to help managers in Baltimore City define tree planting goals and identify priority areas for planting were conducted solely within the city limits. It is the responsibility of the investigators to state what the boundaries of their study units are and what the criteria for inclusion in the urban system are [40]. Such considerations suggest a methodological principle for urban ecology:

Principle 12: Definition of the boundaries and content of an urban study system is set by the researchers based on their research questions or the spatial scope of its intended application.

The second methodological concern is how to frame comparisons within urban systems. Because urban systems are hybrid human–natural complexes and because they are heterogeneous, there are many possible axes of comparison within them [27]. Comparisons may be based on criteria as divergent as the proportion of impervious surface, the amount of woody vegetation, the density of the human population, the surface area of buildings, the accumulated wealth of residents, the lifestyle characteristics of neighborhoods, or the power relationships of local institutions. Many other criteria are possible. However, different criteria are appropriate bases for different comparisons. For example, hydrology and flood risk may be compared on the basis of impervious surface [25]. The amount of woody vegetation may be used in comparisons of the exchange of heat or carbon dioxide between the surface and the atmosphere. The density of human population may be used as a predictor of water quality. Lifestyle contrasts may form the basis for comparisons of environmental decisions made by households in different parts of a metropolis, or in different metropolitan areas [41].

The fact that these many criteria are often patchily arranged suggests that comparisons will rarely be accomplished by running linear transects across metropolitan maps. Although raw distance is

sometimes useful in making coarse-scale urban comparisons, as it was along the urban-rural transect in metropolitan New York [42], in many cases the spatial heterogeneity as it exists on the ground may better be presented as a rearranged, abstract gradient [43]. An example of a gradient abstracted from an array of plots is the gradient of wealth along which bird biodiversity was distributed in Phoenix. Thus, while linear transects may be appropriate bases for initial comparison, ordinated gradients are in fact the general case which should be used to frame comparisons over different scales. The principle is as follows:

Principle 13: Urban comparisons can be framed as linear transects or as abstract gradients, and the abstract comparisons acknowledge the spatial complexity of urban heterogeneity.

The baker's dozen of principles present above frame the current state of urban ecology. As a biological discipline closely related to social and economic sciences and the theory and practice of urban design and planning, it is a relatively new and outward-looking field [13]. Hence, these principles are not the final roster that the field may exhibit. Nor are they in what is likely to be their final form. However, these broad ideas should continue to be useful over the next decade as this field consolidates by accumulating more data on specific urban places through time and by comparing amongst urban areas across the globe. These working principles identify assumptions and suggest hypotheses that can be used to guide future research on urban socio-ecological systems.

Future Directions

Although, as shown above, urban ecology has deep roots as a biological science as well as important parallels in social sciences and the design professions, it remains a young discipline. It is poised for significant growth and increased practical importance. There are three realms in which future directions cluster.

Interdisciplinary Integration for Understanding

Urban areas have been studied by many different disciplines, ranging from sociology, physical and human geography, economics, anthropology, and the more recent offshoots of these classic fields, such as political ecology, political economy, and human ecology. There

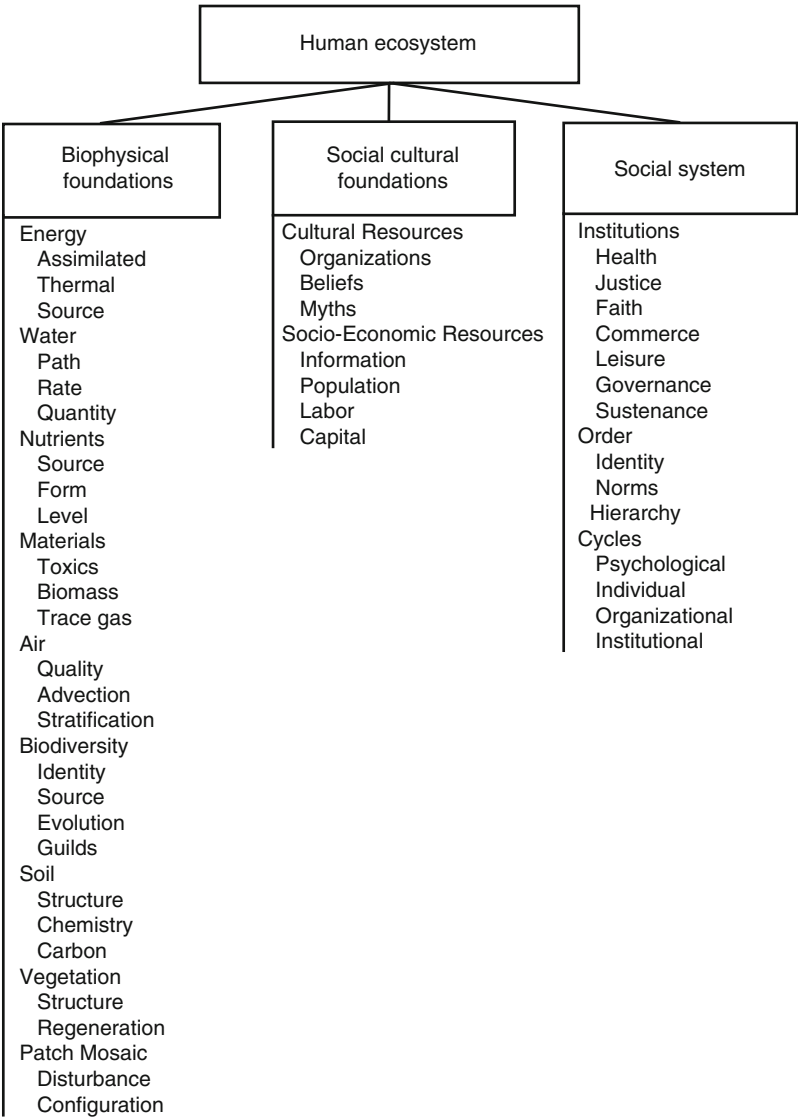
is a large opportunity for integration across these perspectives. Accomplishing that integration is beyond the scope of this article, but ecological science has an important role to play in promoting integration among the diverse perspectives these disciplines represent. Ecology can play this role because it is preeminently an integrative discipline. It incorporates several contrasting perspectives that together define a goal for a more complete understanding of systems that contain biophysical components and biologically driven processes among their complex structures and functions. Ecology as a science focuses on (1) the interactions of organisms and environment through feedbacks, (2) the interactions of organisms with each other, (3) and the transformations that organisms generate within the environment. It is thus a science of interaction, feedback, and change. These three features are among the most important characteristics of urban systems.

Ecology can be a further stimulus to integration within urban systems because of its growing appreciation and study of the role of humans in ecosystems ranging from remote wilderness to suburbs and downtown districts [12]. This biological science has begun to link effectively with economics, with social sciences, with hydrology, and with atmospheric sciences, for example, in order to contribute to more comprehensive models of inhabited and built landscapes [44]. Yet there is still more work to be done on these various frontiers of integration, and the range of urban systems and ecosystem patches within complex city-suburban-exurban systems is an important arena in which to test hypotheses and models combining the perspectives of the different disciplines. Furthermore, data that integrate different disciplinary perspectives and the attempt to generalize cautiously from these data are important empirical activities.

A useful guide to these integrative approaches is the human ecosystem framework. Introduced originally by the social scientist William R. Burch, Jr., and his colleagues [21], this framework has been modified to reflect more bioecological content and thus to better support the integrative program of urban ecology [44]. The framework is not itself a model, but rather an organized roster of potential causes, mechanisms, and interactions upon which specific models can draw. The human ecosystem framework does take a human-centered perspective on inhabited and

managed systems and thus breaks the concept of human ecosystem down into the social system as the primary focus and into the biophysical foundations and sociocultural foundations of that social system (Fig. 1). This framework is an excellent abstract picture of any urban ecosystem. The components contained in this causal framework will have specific instances and representations in any urban area.

The biophysical foundations start metaphorically with the earth, air, fire (or energy), and water perspectives that must comprise any environment [45]. However, to prevent neglecting important environmental features that ecologists have come to appreciate in all systems they study, it goes on to include nutrients, which at some concentration often become pollutants, toxic materials and contaminants, soil structure and



Urban Ecology. Figure 1

The human ecosystem framework as a hierarchy of potential causes and interactions within human ecosystems (Modified after Machlis et al. [21])

chemistry, vegetation composition and dynamics through time, and the spatial mosaic that the biophysical and built components jointly define in urban systems. Each of these categories has at least as much component complexity, if not more, than appears in Fig. 1. The sociocultural foundations include the resources provided by culture in its material, mental, and spiritual forms. Socioeconomic resources among the sociocultural foundation include information, population, labor, and capital. Capital of course refers not only to financial resources but also to the talents and skills of individual persons and the social capital embodied in networks of human interaction. The social system is the structure that enables and constrains the social interactions that all humans depend upon for their immediate well-being. The social system is characterized by the intuitions, in the broad sense, that humans construct to accomplish the tasks of survival, interaction, health, and control. The social system also establishes order among persons and does so in the form of social identity, formal and informal norms of behavior, and the establishment of social rank hierarchies along several dimensions (Fig. 1). The framework identifies cycles as a way of recalling that the components of the social system are not fixed in time or space, but can change with the collective changes in individual and household life cycles, individual and group psychological changes, organizational age as expressed in capacity and flexibility, and the degree of persistence of the various institutional structures in a society. The current intensification of globalization suggests that in many cases local or regional dynamics will be affected by biophysical, economic, and social changes that originate at a distance. For example, the massive migrations from countryside to cities, or from countries in which unemployment is high to different countries where people know or perceive there to be greater opportunities, or the displacements by war and disaster, all suggest looking beyond immediate boundaries for causes of change in city-suburban-exurban systems.

Urban Comparisons

Some generalizations have emerged in discussing the principles. For example, biotic homogenization is a common feature of urban systems [46], although

not all taxa or functional groups express homogenization. Likewise, there is an urban stream syndrome, in which downcutting of stream beds is associated with stranding of flood plains and reduction of riparian and in-stream ecological function, and with simplification of the heterogeneity and biotic diversity in stream channels [47]. On a more hypothetical level are generalizations about the increasing similarity of carbon budgets of cities based on planting and maintenance of mesic vegetation in all types of climates [48]. Although these generalizations seem robust so far, they must be challenged, refined, or rejected based on further comparisons with additional cities. Of course, to conduct appropriate comparisons, cities may have to be more explicitly classified according to some globally relevant schema than they are now [27]. City-suburban-exurban systems differ based on location relative to rivers, streams, or coastal waters; regional climate and form of original vegetation; kind of geological substrate; and exposure to natural hazards. In the broad social realm, cities experience different forms of governance; different histories with respect to development, industry, shipping, and ground transportation; access to commodities and sources of wealth; cultural context and diversity; and porosity of social groups. Not all of these may be major axes of comparison, but they point to some of the many dimensions that may affect differences among cities in their bioecological structure and function.

An important aspect of the comparison of cities is their ability to represent global change processes. Cities may well stand in as a laboratory for global change, as temperate cities are now often drier than their mesic surroundings, and due to the urban heat island effect, are generally hotter. Portions of some highly irrigated desert cities may be an exception and in fact be cooler during some hours than the surrounding arid lands due to the evaporation of massive quantities of surface irrigation water or introduced mesic plants [22]. Another feature of global change that some cities may mimic is a great exposure to flood risk due to the building of floodwalls and levees upstream.

Comparisons within cities can benefit from new land cover classifications. The usual "Anderson Level II" classification uses such categories as commercial, industrial, transportation, and residential in low,

medium, and high intensities of occupancy. These categories may be too coarse for some desirable ecological comparisons [49]. Likewise, the census geography of block groups based on clusters of approximately 400 households may be rather coarse relative to some levels of ecological function [19]. Both the land use/land cover and the census geographies fail, for example, to adequately match the fine-scale watershed behaviors in urban areas. Social perceptions of edges, enclaves, corridors, and so on may also match other scales of observation than the classical tools of differentiation within urban areas. One example of a classification scheme devised to expose aspects of joint biophysical and social differentiation is the High Ecological Resolution Classification for Urban Landscapes and Environmental Systems (HERCULES) [49]. Rather than assuming land use as the determining criterion, it focuses on land cover in order to provide a structural base to test against bioecological functioning. HERCULES classifies patches in urban systems based on (1) surface characteristics, whether bare or paved, (2) presence and amount of either tall or short vegetation, and (3) cover and kinds of buildings. The classes may be extracted from continuous variation along these three axes, or may be defined to represent percentiles of the different categories. This and other classifications are a frontier tool for urban ecology to promote within and cross-system comparisons.

Integration for Practice

Urban ecological science can be integrated with a number of professional practices. This section illustrates its potential for linkage with urban design and planning, environmental justice, and adaptive processes supporting sustainability. Further integration can emerge from dialog with the concerns of urban policy makers and managers.

Integrating with Urban Design and Planning

Urban design incorporates the activities of architects and landscape architects and often focuses on individual sites, on neighborhoods, or spatially extensive development or redevelopment projects. Urban planning, in contrast, brings aspects of that same knowledge set together with knowledge about policy, governance, and regulation and generates coarse-scale designs for large

areas or regions. Master planning is a synonym for this coarser-scale pursuit. Both have traditions that express concern with ecology as a relevant knowledge base or approach. One tradition takes analogies with evolution or ecological processes as a touchstone. This tradition began with Geddes in Scotland, who is one of the founders of modern landscape architecture [50]. Ecological analogies have been common among those who wish to bring an ecological perspective to the urban design and planning professions. This is in part because there has been until recently so little empirical ecological research in urban systems. In this knowledge gap, idealized successional trends from simple to spatially complex systems, ending in a stable “climax,” have been brought to bear. Other analogies adopted in the design professions have been the classical ecological expectation of equilibrium in systems. A still larger and deeply flawed analogy has been the adoption of an organism-like life cycle for cities and towns. The life cycle analogy was used to justify policies of physical intervention to circumvent undesired changes in cities. Of course, the assumption that physical change was sufficient to address social problems is another questionable assumption, often tacit, that was used along with life cycle analysis to justify such interventions as slum clearance. Unfortunately, these analogies are not supported by the contemporary knowledge base in ecological science itself. Succession is a probabilistic process that does not lead inexorably toward some ideal state. Equilibrium composition is rarely obtained in ecological systems, even though material balances and thermodynamic processes can be used to understand the structure and function of ecosystems.

In contrast to the use of ecological analogies, which is necessarily limited by the superficiality of the visualizations transferred to design, ecological substance has informed a growing number of designers. In Europe, analyses of “ecotopes,” patches that hosted biodiversity or complex bioecological structure, were incorporated into plans to meet aesthetic or functional needs the patches were perceived to satisfy. In the realm of green-field development, McHarg [51] famously introduced a multilayered strategy to analyze ecological, physical, historical, and cultural features of an area to be developed. The site design was spatially arranged to optimize the joint benefits and reduce the hazards associated with the layers representing different features of the

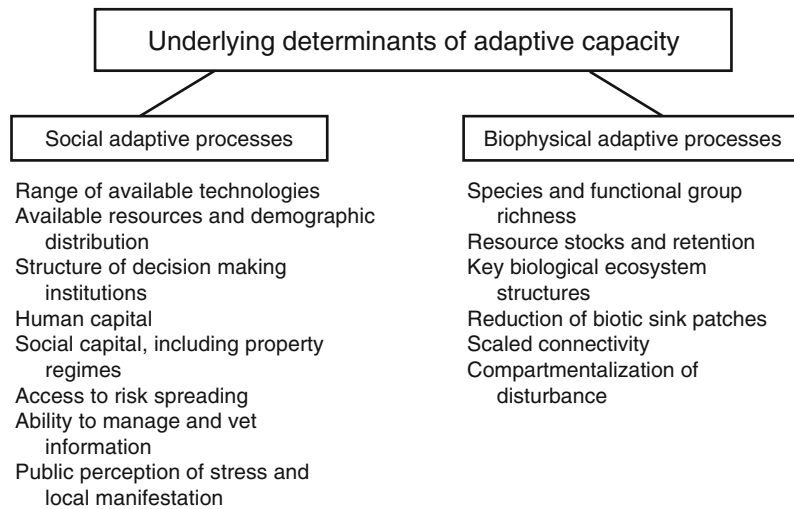
landscape. Spirn [32] applied a careful analytic approach to established urban areas in her work in Boston and west Philadelphia. She used an understanding of watershed dynamics to explain where disempowered communities faced risks due to design in their neighborhood that had been insensitive to ecological processes. This information then informed appropriate neighborhood redevelopment and plans for renewed green space in these old city neighborhoods. This evolving tradition is notable for taking account of both bioecological processes and social concerns. Other designers and planners are promoting the integration of ecological substance with their professional theory and practice [52]. Shared work between designers and ecologists has several advantages. One advantage to working with designers is the entry they provide into the intricacies of development and construction in urban environments, thus making new sites available to their ecological research partners [31]. A second advantage of the increasing interaction between ecologists and designers is to treat constructed design projects as experimental venues [30]. Indeed, because designs provide alternative arrangements of various kinds of surfaces, vegetation, and slopes, for example, they may yield important and novel information on the relationship between urban form and its ecological function [53]. Scaling these relationships from individual small sites to large developments provides another dimension of experimental comparison. Measurements taken in contrasting built projects and landscapes provide an essentially untapped resource for increasing the ecological understanding of urban systems [54].

Environmental Justice Environmental justice refers to the concern about equal access of all persons to environmental goods and services, and equal avoidance of or capacity to respond to environmental hazards. As a social movement, it emerged from communities of color and from communities with poor access to resources and power, in which people noted an unusual association with toxic sites or other serious environmental disamenities. As a result of this origin, environmental justice has an activist dimension and a scholarly dimension. Ecology can contribute to the scholarship of environmental justice by enhancing the understanding of environmental benefits and environmental

hazards. Traditionally, environmental justice has focused on environmental negatives or disamenities. In cities, ecologists and related scientists are increasing the stock of knowledge about not only the disamenities, but increasingly about the benefits that emerge from the biotic components of human ecosystems. This focus on benefits is part of the study of ecosystem services [55]. Examples of urban ecosystem services include the mediation of local climate by trees, the tempering of human behavior by green views, the reduction of stormwater flows, provision of recreation, conversion of some pollutants to neutral forms, and the sequestration of carbon dioxide in biomass, soils, and organic litter [56]. Cultural benefits also accrue to the biotic components of urban ecosystems. These include a sense of place provided by ecological structures such as particular patches of vegetation or streams. Environmental justice can be enhanced by a broader understanding of the spatial distribution of hazards such as fire and exposure to winds, and to the potential for mitigation and other positive services provided by the biotic components of urban ecosystems.

Sustainability and Adaptive Processes An emerging frontier in urban systems is the nature and processes of sustainability. Sustainability is defined in two main ways. One definition suggests that sustainability is the capacity to meet present needs without compromising the ability of future generations to meet their own needs. A second definition focuses on the reconciliation of demands within economic, social, and environmental spheres that permit the entire integrated system to persist and continue to adapt to change. This second definition is important for giving equal weight to the three facets of the global hybrid system as well as accommodating social justice as a key aspect of the social dimension. Indeed, some scholars define sustainability in terms of economy, environment, and equity, following the assumption that a social system that does not provide for equity is less likely to persist.

Sustainability, as the ability to adjust to changing conditions originating within and beyond the system of interest, is closely related to the concept of resilience. Resilience is the capacity of systems to experience disturbance and shocks and still remain within a given structural or functional domain. Such resilience requires adaptive processes (Fig. 2). Adaptive processes



Urban Ecology. Figure 2

Underlying determinants of adaptive capacity in socio-ecological systems. Social components derived from Yohe and Tol [57] and biophysical components derived from Gunderson and Pritchard [58] and Walker et al. [59]

are both social [57] and bioecological [58, 59], in keeping with the understanding of most contemporary ecosystems as human ecosystems (Fig. 2). Adaptive processes bring the components of human ecosystems into the dynamic realm of sustainability and resilience.

Sustainability has become a goal of many urban jurisdictions. This development is the latest stage in an ongoing evolution of cities in industrial countries. The industrial city was the epitome of overcrowding, contaminated water, and industrially generated pollution, for example. Reformers followed two paths to correct the flaws of the industrial city. One was to abandon it altogether. This strategy motivated the establishment of suburbs, or of garden cities in green sites well outside the shadow of the dark Dickensian industrial city. The second strategy was to alter the industrial city itself. Providing clean water, removing waste, alleviating overcrowding, and reducing the local impact of pollution were hallmarks of this second strategy, which can be labeled, following historian Martin Melosi [60], the “sanitary city.” The sanitary city installed infrastructure to carry sewage away from the populated areas. Ultimately, treatment was added, so that downstream systems did not have to bear the burden of the displaced sewage. Smokestacks were built taller so that local pollution was alleviated. Ultimately, newer fuels or scrubbing technologies for coal-fired

installations also contributed to reduced pollution. Sanitary landfills were established, and odoriferous wetlands were drained or filled. Many of the solutions employed in the sanitary city were engineered tactics to deal with specific problems. The departmental structure of cities followed this issue-by-issue approach, so that water supply, waste water, solid waste, streets, housing, parks, and so on were dealt with by separate agencies of municipal government.

The emerging sustainable city contrasts with the key characteristics of the older sanitary city. Cities will always be heterotrophic systems, garnering food and other resources from beyond their boundaries. Furthermore, they will generate waste that must be dealt with, requiring commitments of land and other resources – an urban footprint – that extends well beyond their geopolitical boundaries [61]. However, as an ideal, sustainable cities aim to reduce their ecological footprint. A sustainable city is one that reduces its demand and its impact on the environment by reducing resource use and waste generation. Sustainable cities manage potable water, gray water, and stormwater in ways that reduce flood risks, contamination of water, and waste of potable water. They reduce energy use via building standards, landscape design, and reducing per capita use of fuel in transportation. Limiting sprawl while maintaining space for ecological processes and

capacities in the urban matrix is an important feature of the sustainable city. Hence, other dimensions of sustainability must be supported. In the environmental realm, reducing vulnerability to natural and human-generated hazards is a sustainability goal, as are supporting native biodiversity and achieving the ecosystem services that emanate from green spaces in the urban matrix. Reduced contribution to global warming can be achieved through enhanced carbon sequestration.

Sustainability by definition also must deal with social processes. Too often in the past, urban interventions have focused on the physical form of the city, assuming that social benefits will follow. Contemporary focus on sustainability recognizes that social capital, community cohesion, and social equity are issues that must be addressed in and of themselves for generating and maintaining liveable cities. Equity is a major component of the social realm, as discussed above, and is reflected in exposure to environmental hazards, access to environmental benefits, and inclusion in the process of making environmentally relevant decisions.

The economic aspect of sustainability is well known, and because of the traditional predominance of attention in urban studies to issues of jobs, finance, the real estate industry, and the relationships of business and political processes, it is not necessary to review it further in this article. Clearly, however, economic resources and sustainability of economic capacity contribute to well-being in urban systems.

One of the potentially most powerful aspects of the sustainable city ideal is how management and policy are conducted in this new kind of city. As mentioned before, the sanitary city is managed by discrete departments or authorities, often having little cross-communication. For example, the concern with street cleaning are isolated in the transportation department, whereas street sweeping in fact affects stormwater quality. Another cross-cutting issue is how the presence of vegetation can influence aggression, which joins the concerns of a parks department with those of public safety. These simple examples suggest that sustainability can be enhanced by integrated management in urban systems. Indeed, cross-cutting management and policy are hallmarks of the emerging sustainable city. This suggests that sustainability can be usefully considered a central function of city administration and not an isolated pursuit to be marginalized in a special office isolated from

the traditional activities of municipal government. Indeed, there are growing numbers of examples of cross-sectoral management as part of sustainability strategies. For example, in the city of Baltimore, neighborhood tree planting and gardening, removal of unneeded pavement in school yards, altered street cleaning schedules, community environmental activities, and installation of fine-scale best management practices such as rain gardens specifically designed to be both attractive and to allow infiltration of stormwater into the soil, engage multiple city departments in specific neighborhoods to achieve social, environmental, and economic revitalization. Ecological research can help inform and evaluate sustainability plans, which are often mostly metaphorical or assessed only by indices of human outcomes.

Communication between researchers and decision-makers is a key ingredient in the success of the sustainable city. Growing experience shows that communication is enhanced when approached as a two-way dialog, rather than a one-way flow of information from science to policy. Indeed, such a dialog can help shape scientific research and identify practical management projects that provide data about the structure and function of urban ecosystems. Mutual respect for the knowledge, constraints, and rewards of each group involved in the dialog is a firm foundation for effectively shaping the science and the practice of urban sustainability. Establishing effective and lasting forums for such communication is a practical part of the frontier of urban ecological science. Such dialog is one tool for the sustainable city, which remains a goal rather than a reality. Combining the knowledge, concerns, and networks represented by problem solving and knowledge generation is perhaps the most urgent frontier for further research and action.

Bibliography

Primary Literature

1. Collins JP et al (2000) New urban ecology. *Am Sci* 88:416–425
2. Elvidge CD et al (2004) US constructed area approaches the size of Ohio. *Euro Opt Soc* 85:233–240
3. Grimm NB et al (2000) Integrated approaches to long-term studies of urban ecological systems. *BioScience* 50:571–584
4. Numata M (1977) The impact of urbanization on vegetation in Japan. In: Miyawaki A, Tuxen R (eds) *Vegetation science and environmental protection*. Maruzen, Tokyo

5. Loucks OL (1994) Sustainability in urban ecosystems: beyond an object of study. In: Platt RH, Rowntree RA, Muick PC (eds) *The ecological city: preserving and restoring urban biodiversity*. University of Massachusetts Press, Amherst, pp 48–65
6. Alberti M et al (2003) Integrating humans into ecology: opportunities and challenges for studying urban ecosystems. *BioScience* 53:1169–1179
7. Pickett STA, Grove JM (2009) Urban ecosystems: what would Tansley do? *Urban Ecosyst* 12:1–8
8. Sukopp H (1990) Urban ecology and its application in Europe. In: Sukopp H, Hejny S, Kowarik I (eds) *Urban ecology: plants and plant communities in urban environments*. SPB Academic, The Hague, pp 1–22
9. Stearns F, Montag T (eds) (1974) *The urban ecosystem: a holistic approach*. Dowden, Hutchinson and Ross, Stroudsburg, p 217
10. Wu J, Loucks OL (1995) From balance of nature to hierarchical patch dynamics: a paradigm shift in ecology. *Q Rev Biol* 70:439–466
11. Boyden S et al (1981) *The ecology of a city and its people: the case of Hong Kong*. Australian National University Press, Canberra, pp 21–473
12. Redman C, Grove JM, Kuby L (2004) Integrating social science into the long-term ecological research (LTER) network: social dimensions of ecological change and ecological dimensions of social change. *Ecosystems* 7:161–171
13. Grimm NB et al (2008) Global change and the ecology of cities. *Science* 319:756–760
14. Fernández-Juricic E, Jokimäki J (2001) A habitat island approach to conserving birds in urban landscapes: case studies from southern and northern Europe. *Biodivers Conserv* 10:2023–2043
15. Gibb H, Hochuli DF (2002) Habitat fragmentation in an urban environment: large and small fragments support different arthropod assemblages. *Biol Conserv* 106:91–100
16. Crooks KR et al (2001) Extinction and colonization of birds on habitat islands. *Conserv Biol* 15:159–172
17. Leston LFV, Rodewald AD (2006) Are urban forests ecological traps for understory birds? An examination using northern cardinals. *Biol Conserv* 131:566–574
18. Groffman PM et al (2003) Down by the riverside: urban riparian ecology. *Front Ecol Environ* 1:315–321
19. Cadenasso ML et al (2008) Exchanges across land-water-landscape boundaries in urban systems: strategies for reducing nitrate pollution. *Ann NY Acad Sci* 1134:213–232
20. Light JS (2009) *The nature of cities: ecological visions and the American urban professions 1920–1960*. Johns Hopkins University Press, Baltimore
21. Machlis GE, Force JE, Burch WR (1997) The human ecosystem. 1. The human ecosystem as an organizing concept in ecosystem management. *Soc Nat Resour* 10(4):347–367
22. Baker LA et al (2002) Urbanization and warming of Phoenix (Arizona, USA): impacts, feedbacks and mitigation. *Urban Ecosyst* 6:183–203
23. Troy AR et al (2007) Predicting opportunities for greening and patterns of vegetation on private urban lands. *Environ Manage* 40:394–412
24. Alig RJ, Kline JD, Lichtenstein M (2004) Urbanization on the US landscape: looking ahead in the 21st century. *Landscape Urban Plann* 69:219–234
25. Gurnell A, Lee M, Souch C (2007) Urban rivers: hydrology, geomorphology, ecology and opportunities for change. *Geogr Compass* 1(5):1118–1137
26. Snep RPH et al (2006) How peri-urban areas can strengthen animal populations within cities: a modeling approach. *Biol Conserv* 127:345–355
27. Dow K (2000) Social dimensions of gradients in urban ecosystems. *Urban Ecosyst* 4:255–275
28. Theobald DM (2004) Placing exurban land-use change in a human modification framework. *Front Ecol Environ* 2:139–144
29. Radeloff VC et al (2005) The wildland urban interface in the United States. *Ecol Appl* 15:799–805
30. Outen DC (2007) Pioneer on the frontier of smart growth: the Baltimore County, MD experience, in *Smart growth @ 10: a critical analysis of Maryland's landmark landuse program*. The National Center for Smart Growth Research and Education, Annapolis and College Park, Maryland, p 49
31. Felson AJ, Pickett STA (2005) Designed experiments: new approaches to studying urban ecosystems. *Front Ecol Environ* 3:549–556
32. Spirn AW (2005) Restoring Mill Creek: landscape literacy, environmental justice and city planning and design. *Landscape Res* 30:395–413
33. Groffman PM et al (2002) Soil nitrogen cycling processes in urban riparian zones. *Environ Sci Technol* 36:4547–4552
34. Kennedy C, Cuddihy J, Engel-Yan J (2007) The changing metabolism of cities. *J Ind Ecol* 11:43–59
35. McKinney ML (2002) Urbanization, biodiversity, and conservation. *BioScience* 52:883–891
36. Beissinger SR, Osborne DR (1982) Effects of urbanization on avian community organization. *Condor* 84:75–83
37. Yeh PJ (2004) Rapid evolution of a sexually selected trait following population establishment in a novel habitat. *Evolution* 58:166–174
38. Slabbekoorn H, Peet M (2003) Birds sing at a higher pitch in urban noise. *Nature* 424:267
39. Wyckoff PH, Webb SL (1996) Understory influence of the invasive Norway maple (*Acer platanoides*). *Bull Torrey Bot Club* 123:197–205
40. McIntyre NE, Knowles-Yanez K, Hope D (2000) Urban ecology as an interdisciplinary field: differences in the use of “urban” between the social and natural sciences. *Urban Ecosyst* 4:5–24
41. Zhou W et al (2008) Can money buy green: demographic and socioeconomic predictors of lawncare expenditures and lawn greenness in urban residential areas. *Soc Nat Resour* 22:744–760
42. Pouyat RV, Turechek WW (2001) Short- and long-term effects of site factors on net N-mineralization and nitrification rates along an urban-rural gradient. *Urban Ecosyst* 5:159–178
43. McDonnell MJ et al (1997) Ecosystem processes along an urban-to-rural gradient. *Urban Ecosyst* 1:21–36

44. Pickett STA et al (2011) Urban ecological systems: scientific foundations and a decade of progress. *J Environ Manag* 92:331–362
45. Spirn AW (1984) *The granite garden: urban nature and human design*. Basic Books, New York, pp 1–334
46. Clergeau P et al (2006) Avifauna homogenisation by urbanisation: analysis at different European latitudes. *Biol Conserv* 127:336–344
47. Walsh CJ et al (2005) The urban stream syndrome: current knowledge and the search for a cure. *J N Am Benthol Soc* 24:706–723
48. Carreiro MM et al (2009) Carbon and nitrogen cycling in soils of remnant forests along urban-rural gradients: case studies in the New York metropolitan area and Louisville, Kentucky. In: McDonnell MJ, Hahs A, Breuste J (eds) *Ecology of cities and towns: a comparative approach*. Cambridge University Press, New York, pp 308–328
49. Cadenasso ML, Pickett STA, Schwarz K (2007) Spatial heterogeneity in urban ecosystems: reconceptualizing land cover and a framework for classification. *Front Ecol Environ* 5:80–88
50. Welter VM (2002) *Biopolis: Patrick Geddes and the city of life*. MIT Press, Cambridge, p 355
51. McHarg I (1969) *Design with nature*. Doubleday/Natural History Press, Garden City, p 197
52. Musacchio LR (2009) Pattern:process metaphors for metropolitan landscapes. In: McDonnell MJ, Hahs A, Breuste J (eds) *Ecology of cities and towns: a comparative approach*. Cambridge University Press, New York, pp 484–502
53. Pickett STA, Cadenasso ML (2008) Linking ecological and built components of urban mosaics: an open cycle of ecological design. *J Ecol* 96:8–12
54. Cook WM et al (2004) Learning to roll with the punches: adaptive experimentation in human dominated systems. *Front Ecol Environ* 2:467–474
55. Boone CG et al (2009) Parks and people: an environmental justice inquiry in Baltimore, Maryland. *Ann Assoc Am Geogr* 99:767–787
56. Bolund P, Hunhammar S (1999) Ecosystem services in urban areas. *Ecol Econ* 29:293–301
57. Yohe G, Tol RSJ (2002) Indicators for social and economic coping capacity – moving toward a working definition of adaptive capacity. *Glob Environ Change* 12:25–40
58. Gunderson LH, Pritchard L Jr (eds) (2002) *Resilience and the behavior of large-scale systems*. Island Press, Washington, DC, p 287
59. Walker B et al (2004) Resilience, adaptability and transformability in social-ecological systems. *Ecol Soc* 9(2): Article 5
60. Melosi MV (2000) *The sanitary city: urban infrastructure in America from colonial times to the present*. Johns Hopkins University Press, Baltimore, 600 pp
61. Rees WE (1992) Ecological footprints and appropriated carrying capacity: what urban economics leaves out. *Environ Urbanization* 4:121–129
- Beatley T (2000) *Green urbanism: learning from European cities*. Island, Washington, DC
- Bullard RD (ed) (2005) *The quest for environmental justice: human rights and the politics of pollution*. Sierra Club Books, San Francisco
- Burch WR Jr (1988) Human ecology and environmental management. In: Agee JK, Johnson DR (eds) *Ecosystem management for parks and wilderness*. University of Washington Press, Seattle, pp 145–159
- Burger J (1999) *Animals in towns and cities*. Kendall/Hunt Publishing Company, Dubuque
- Cadenasso ML, Pickett STA (2008) Urban principles for ecological landscape design and management: scientific fundamentals. *Cities Environ* 1(2):1–16
- Catterall CP (2009) Responses of faunal assemblages to urbanisation: global research paradigms and an avian case study. In: McDonnell MJ, Hahs A, Breuste J (eds) *Ecology of cities and towns: a comparative approach*. Cambridge University Press, New York, pp 129–155
- Curwell S, Deakin M, Symes M (eds) (2005) *Sustainable urban development, volume 1: the framework and protocols for environmental assessment*. Routledge, New York
- Forman RTT (2008) *Urban regions: ecology and planning beyond the city*. Cambridge University Press, New York
- Frumkin H, Frank L, Jackson RB (2004) *Urban sprawl and public health: designing, planning and building for healthy communities*. Island, Washington, DC
- Gleick PH (2009) *The world's water: 2008–2009*. Island, Washington, DC
- Grove JM (2009) Cities: managing densely settled social-ecological systems. In: Chapin FS III, Kofinas GP, Folke C (eds) *Principles of ecosystem stewardship: resilience-based natural resource management in a changing world*. Springer, New York, pp 281–294
- Gunderson LH (2000) Ecological resilience – in theory and application. *Annu Rev Ecol Syst* 31:425–439
- Marzluff JM, Schulenberger E, Endlicher W, Alberti M, Bradley G, Ryan C, Simon U, ZumBrunne C (eds) (2008) *Urban ecology: an international perspective on the interaction between humans and nature*. Springer, New York
- McDonnell MJ, Breuste J, Hahs A (2009) Introduction: scope of the book and need for developing a comparative approach to the ecology of cities and towns. In: McDonnell MJ, Hahs A, Breuste J (eds) *Ecology of cities and towns: a comparative approach*. Cambridge University Press, New York, pp 1–5
- McGrath BP, Marshall V, Cadenasso ML, Grove JM, Pickett STA, Plunz R, Towers J (eds) (2007) *Designing patch dynamics*. Columbia University Graduate School of Architecture, Preservation and Planning, New York
- Sukopp H, Numata M, Huber A (1995) *Urban ecology as the basis of urban planning*. SPB Academic, The Hague
- United Nations Population Fund (2007) *State of world population 2007: unleashing the potential of urban growth*. United Nations Population Fund, New York

Books and Reviews

- Alberti M (2008) *Advances in urban ecology: integrating humans and ecological processes in urban ecosystems*. Springer, New York

Urban Forest Function, Design and Management

KJELL NILSSON¹, CECIL C. KONIJNENDIJK¹,
ANDERS BUSSE NIELSEN²

¹Danish Centre for Forest, Landscape and Planning,
University of Copenhagen, Frederiksberg C, Denmark

²Department of Landscape Management, Design and
Construction, Swedish University of Agricultural
Sciences, Alnarp, Sweden

Article Outline

Glossary

Definition of the Subject with Brief Historical
Background

Introduction: Urban Forest Resources in a Global
Context

The Benefits of Urban Greening

Urban Forest Policies

Design and Management of the Urban Forest

Future Directions

Bibliography

Glossary

Green infrastructure Green infrastructure is a planned network of multifunctional green spaces and interconnecting links which is designed, developed and managed to meet the environmental, social, and economic needs of communities.

Urban forest The sum of all woodland and other tree-based vegetation in and immediately surrounding an urban area.

Urban forestry The art, science, and technology of managing trees and forest resources in and around urban communities for the environmental, social, economic, and esthetic benefits trees provide to people.

Urban green space All vegetated lands in urban areas, including parks, woodland, gardens, cemeteries, orchards, etc. Often urban green space is used in terms of publicly owned green space.

Urban greening The planning and management of all urban vegetation to create or add values to the local community in an urban area.

Urban woodland Forested land located in or close to urban agglomerations, with a multiple forest functions approach. Typically, social and environmental forest services are the focus of urban woodland management.

Definition of the Subject with Brief Historical Background

Land use planners have increasingly been focusing on urban green infrastructure rather than individual green elements. Moreover, politicians, researchers, and practitioners have had to deal with the contributions of this urban green infrastructure to the quality of urban life and environment. They have started to realize that more integrated green area planning and management are required to meet current societal demands when operating in high-pressure environments. This has led to the emergence of new concepts and approaches of a more integrative kind. Urban greening, for example, has developed as the planning and management of all urban vegetation to create or add values to the local community in an urban area [40, 66]. Another more integrative concept resulting from the above-mentioned developments is that of urban forestry.

Urban forestry has become defined as the art, science, and technology of managing trees and forest resources in and around urban communities for the environmental, social, economic, and esthetic benefits trees provide to people (developed from [27, 50]). In its broadest sense, urban forestry embraces a multi-managerial system that includes municipal watersheds, wildlife habitats, outdoor recreation opportunities, landscape design, recycling of municipal wastes, tree care in general, and the production of wood fiber as a raw material. Urban forestry includes activities carried out in the city center, suburban areas, and the peri-urban or interface area with rural lands.

Growing awareness about the need for more integrated approaches to urban green planning and management, and specific threats to urban tree populations caused by pests and diseases (such as Dutch Elm Disease) led to the emergence of urban forestry in North America during the 1960s. Europe and other parts of the world followed with, for example, international research networking intensifying during the late 1990s (e.g., [35, 39, 50]). In spite of this, urban forestry is an

emerging and still developing field. Differences can be noted in the definition and implementation of the urban forestry concept in different parts of the world, with, for example, a traditional focus on urban woodland in Europe (e.g., [40]). However, the concept of urban forestry as encompassing the planning, design, establishment, and management of trees and forest stands with amenity values situated in or near urban areas, has become more widely accepted.

In Europe, the definition of urban forestry has been adapted into the Urban Forestry Matrix see (Fig. 1).

The matrix shows the integrative scope of urban forestry, incorporating urban and peri-urban tree resources, including single trees and groups of trees along streets, in public and private parks and gardens, cemeteries, and so forth. The distinction between the three types of locations included in the matrix arises from three different levels of planning and management hierarchy, stress, establishment techniques, average life times, and costs in relation to establishment and management [55]. Street trees, for example, are usually single trees with a low average lifetime due to a high

stress level. Moreover, street trees generally involve the highest management costs. Park trees are also to be found individually or in small groups, with a medium or high average lifetime, medium stress level, and medium costs for establishment and management. Urban woodland trees are usually established in stands by means of seeding or planting of small trees, with a high average lifetime, low establishment cost, and, relatively, low management costs.

The matrix also shows the different relationships and interactions between human society and urban forest resources. At the more strategic level, form and function of urban forests provide a basis for policy-making, planning, and design. The selection of plants for the urban environment, and tree establishment involve technical activities. Management, finally, is aimed at bringing strategies and technical operations together in order to maintain and develop sustainable and multifunctional urban forest resources.

As visualized in the matrix, urban forestry is an integrative approach, combining street and garden trees with urban woodland, and extending from

	Street trees	Trees in parks, private yards, cemeteries	Urban woodlands
Form, functions, design, policies and planning			
Selection, establishment and other technical approaches			
Management			

Urban Forest Function, Design and Management. Figure 1
The Urban Forestry Matrix (Based on [37], photos by Kjell Nilsson)

policy-making to establishment, maintenance, and management activities. Integration has been described as one of the key characteristics (and strengths) of urban forestry [35]. Other characteristics include its strategic scope, its social inclusiveness (emphasizing stakeholder involvement), its focus on multifunctionality of the urban forest resource, and its embracing of the urban settings and dynamics in which urban forests are situated.

Introduction: Urban Forest Resources in a Global Context

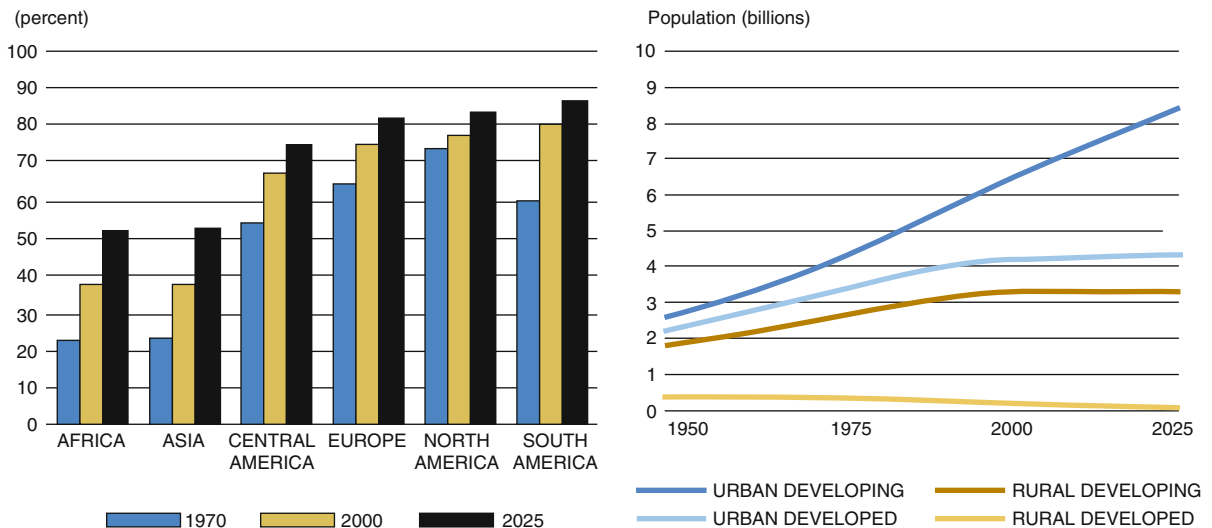
In 2008, for the first time in human history, the number of people living in urban areas exceeded the rural population. Urban areas in developing countries will account for nearly 90% of the projected world population increase of 2,700 million people between 1995 and 2030.

Urbanization is a worldwide phenomenon; the World Resources Institute [87] estimated that in 2025, more than 50% of the African and Asian populations would be living in urban areas. In Central and South America these figures will be between 75% and 85% (Fig. 2).

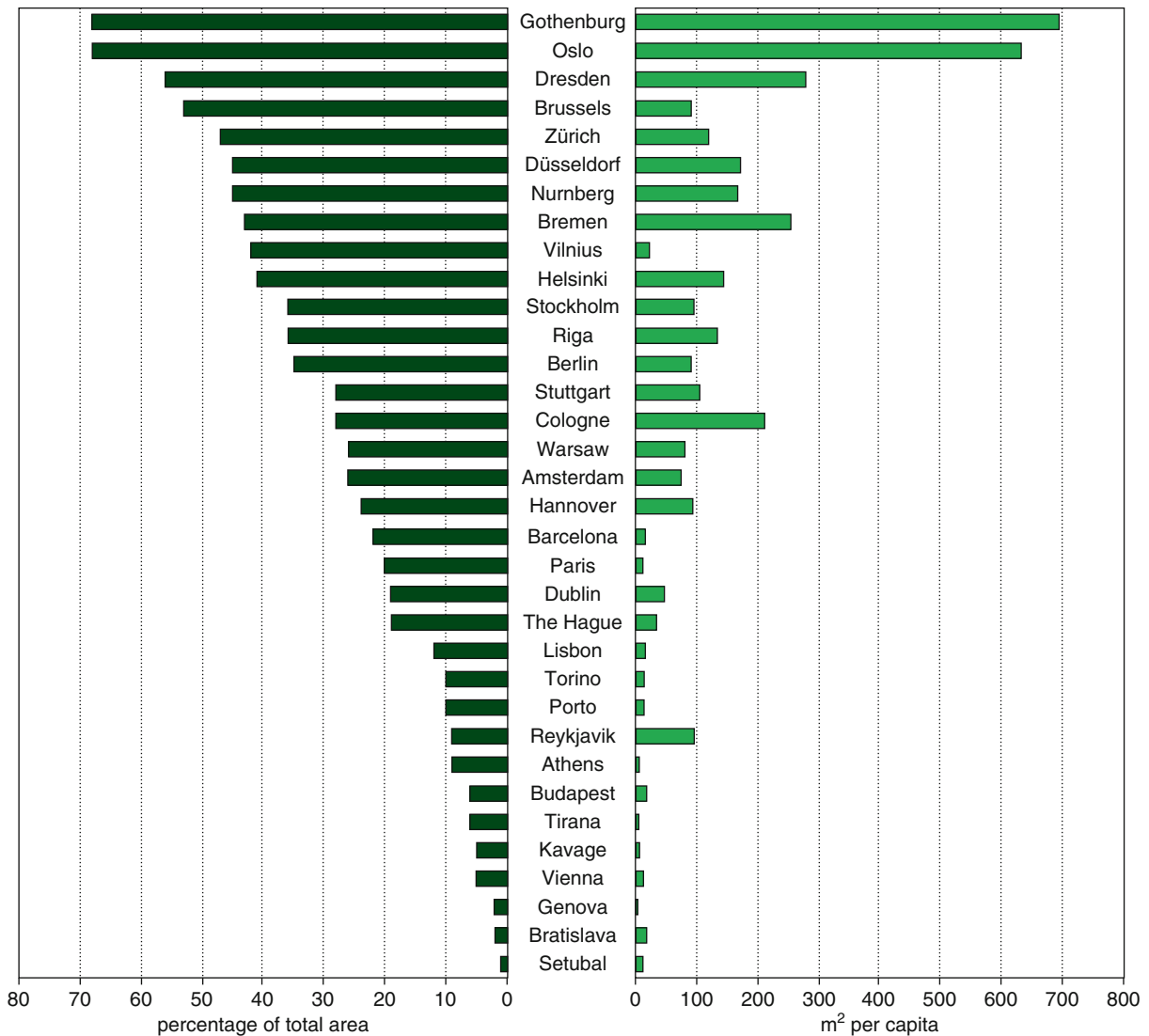
Managing urban population change will be one of the most important challenges during the next decades, along with moderating the impacts of climate change.

In developed countries, the urban future will involve dealing with complex changes in the composition of urban populations and containing urban sprawl beyond the suburbs to retain the critical ecosystem services that will sustain population growth. In developing countries, where 80% of the world's population resides, central issues will be how to cope with an unprecedented increase in the number of people living in urban areas and with the growing concentration of these urbanites in large cities with millions of residents and declining availability of natural resources.

Throughout Europe, the proportion of urban green areas varies considerably, from less than 5% in cities like Tirana and Bratislava to nearly 70% of the area of Oslo and Gothenburg [11] (Fig. 3). In comparison, the figure for Mexico City is only 2.2%; in relation to the number of inhabitants, this only provides 1.94 m² per inhabitant, which is far below the 9 m² recommended by the World Health Organization [6]. However, these figures should be taken cautiously since definitions of green space and city boundaries differ a lot between cities. A suggested measure of urban environmental quality is the location of green areas within a walking distance of 15 min or less from all housing areas. This criterion is met for the citizens in, for example, Brussels, Copenhagen, Glasgow, Madrid, Milan, and Paris and, in general, for more than 50% of the population in most European cities [78].



Urban Forest Function, Design and Management. Figure 2
Estimated urban population by region (From [87])



Urban Forest Function, Design and Management. Figure 3

Green areas in selected European cities [11]

The Benefits of Urban Greening

The green infrastructure of cities provides urban society with an essential range of goods and services, representing the three dimensions of sustainable development: Environmental, economic, and social sustainability.

Environmental and Ecological Functions

Reduce Air Pollution Trees intercept particulate matter and absorb such gaseous pollutants as ozone,

sulfur dioxide, and nitrogen dioxide, thus removing them from the atmosphere. Several studies conducted in the USA as well as in Germany and China have shown that urban tree cover has the potential for improving air quality [7, 57, 88]. Protective plantations along heavily trafficked roads and around industrial areas are, therefore, an effective means of reducing air pollution. In Santiago de Chile, urban forests were found to be a cost-effective air quality improvement policy [15]. But this should obviously not be taken as an excuse for neglecting to combat pollution at its source.

Even though plants absorb carbon dioxide and produce oxygen, it is still important not to assign plants excessive significance to the urban environment. Harris [26] reminds that plants really have only a minor effect on the carbon dioxide and oxygen content of urban air. Photosynthesis in the oceans accounts for between 70% and 90% of the world's total oxygen production, for which reason it is absolutely vital that they be protected against pollution. Even a minor reduction in the oxygen content of the air will cause a large percentage increase in its carbon dioxide content, which would reinforce the greenhouse effect, thus leading to a rise in the global temperature. Harris [26] also stresses that urban vegetation has an especially beneficial effect on air pollution through its ability to cool urban heat lands and reduce the quantity of airborne particles.

Protect Water Resources and Reduce Stormwater Runoff Before modern urban areas made their appearance, rainwater was handled locally. It ran in stone fascines in the ground or out onto fields and pastures. The towns also had streams and wetlands that functioned as recipients. The great change came after the Second World War, when serious attempts were made to drain the water away from residential areas, streets, and squares in underground sewer pipes. Thus, the traditional water cycle of the towns lost its main source. Instead, costly drinking water is now used for watering greens and plantations.

When it falls on roofs and paved areas, rain has become an environmental problem, instead of being a resource that appeals to the senses and benefits plants and animals. Run-off from roads, streets, parking areas, and even some roofs is polluted. Thus, draining rainwater into sewers means that these pollutants end in watercourses, lakes, and the sea. On the other hand, rainwater that has filtered through the soil is considerably cleaner than the water from purification plants [30].

Protection of riparian woodland can be of particular importance for surface water quality. The hydrological function of urban woodland and trees is increasingly stressed as protection of drinking water resources. For example, in Denmark, new woodland areas established close to cities consider this function as a primary one next to recreational benefits [29].

Reduce Harmful Influence of Sun, Wind, and Temperature By transpiring water and shading surfaces, trees lower local air temperatures. Because trees lower air temperatures, shade buildings in the summer, and block winter winds, they can reduce the consumption of building energy and consequently reduce the emission of pollutants from power-generating facilities [48].

Temperatures will change in the future and the so-called urban heat island effect will strengthen the effects of climate change. Results from the ASCCUE (Adaptation Strategies to Climate Change in the Urban Environment) project show that green areas can play an important role for smoothing the effects of rising temperatures. The results show that for the city of Manchester increasing green cover by 10% in urban areas could keep extreme surface temperatures at current levels up until the end of the century, despite climate change [21].

Biodiversity Older gardens and parks often contain noticeably rich biodiversity [59]. These are the main habitats of urban plants and animals. Older, well-established installations attract, for instance, birds and mammals, the natural habitat of which is the forest. Since an increasing part of the population lives in urban areas and receives its daily perception of nature therein, nature in urban areas is important to environmental awareness and an understanding of nature.

Urban landscapes create new habitats for species establishment and germination and colonization sites [86, 89]. These new habitats can affect biodiversity and change community structure. Although an urban landscape may not have a full complement of natural species, the existing natural and seminatural ecosystems can provide a suite of benefits compared to the urban built infrastructure and other technologies.

Economic Functions

The aim of sustainable development is to meet the needs of both present and future generations by promoting development that does not degrade the environment and is technically appropriate, economically viable, and socially acceptable.

Food from trees in private gardens, fruit orchards, or allocated plots in public gardens can contribute

significantly to food security in developing countries [41]. Low-care wild edible plants are often excellent candidates for multipurpose use as ornamental roadside plantings. Older fruits varieties in aging orchards are an important genetic repository.

Woodfuel provides between 25% and 90% of urban household energy supplies; it is particularly important as a source of energy in smaller urban centers in developing countries, especially in dry zones [41]. Poor urban households spend a significant proportion of their cash income in obtaining wood energy. If the urban poor population continues to grow, an increase in the consumption of traded woodfuel is likely to be a consequence. Under favorable circumstances, fuelwood from non-rural forests and agroforestry systems can contribute significantly to fuelwood supply.

In rich countries, urban green spaces can contribute to the local economy, by enhancing real estate prices and by attracting economic activity and improving the local tax base. Several Scandinavian studies show that the prices of apartments are significantly raised in the presence of nearby woodlands or other types of green space (e.g., [81]). Trees and greenspace also help to conserve global energy reserves in developed nations where they are consumed at the greatest rates.

Social Functions

Health Aspects City life is stressful, but research shows that urban green areas have a beneficial influence on the health and well-being of the urban population. Several studies conducted in Europe and the USA indicate that visits to green areas can counteract stress, renew vital energy, and speed healing processes [56].

Kaplan and Kaplan [32] formulated a theory on the interaction between man's attention and his surroundings. This means that urban living, with fast vehicles, flashing neon signs, and strong colors, causes constant stress. The research indicates that vegetation and nature reinforce one's spontaneous attention, allow one's sensory apparatus to relax, and infuse one with fresh energy. Visits to green areas bring relaxation and sharpen one's concentration, since one only needs to use one's spontaneous attention. At the same time, one

gets fresh air and sunlight, which have significance for one's diurnal and annual rhythms.

Furthermore, Ulrich [83] showed that hospitalized patients recovered faster when they had a view through a window, allowing them to see trees. In 1991, he conducted a new study showing a gory film on industrial accidents to 120 people [84]. Half of the people were then shown a nature film, whereas the other half were shown a film on the city, with sequences of buildings and traffic. The subjects' heartbeat, muscular tension, and blood pressure were monitored throughout. All subjects exhibited strong signs of stress during the first film, on industrial accidents. The stress levels of the half of the subjects that then watched the nature film had returned to a normal level after 4–6 min, whereas the half that watched the film on buildings and traffic continued to exhibit high stress levels.

Another important health function is recreation provided by parks in areas that are used for recreational activities such as walking, biking, and spaces for children to play. Finally, when several metric tons of air pollutants are removed by urban forests in Mexico City, this also can lead to substantial health benefits [14].

Educational Functions Closely related with the ecological and social values of urban green spaces is their educational function. Since a rapidly increasing part of the world's population lives in urban areas and receives its daily perception of nature therein, nature in urban areas is important to increase environmental awareness and an understanding of nature.

Richard [45] coined the term *nature deficit disorder*, to refer to the trend that children are spending less time outdoors resulting in a wide range of behavioral problems. He claims that causes for the phenomenon include parental fears for their children in urban areas, restricted access to natural areas around cities, and the lure of the computer and television screen. Result of nature deficit disorder can include childhood obesity, attention disorders, and depression. Providing youth time in nature during the school day can help reduce anxiety and have a positive effect on the attention span, stress reduction, creativity, cognitive development, and children's sense of wonder and connection to the earth.

Urban Forest Policies

Urban Forest Policies Defined

A policy is a generally agreed-to and purposeful course of action that has important consequences for a large number of people and for a significant number and magnitude of resources [12]. Policies can be informal as well as formalized in written documents. Usually policies are selected from a number of alternative courses of action – by government agencies, as well as by institutions, groups, and even individuals.

Drawing on the political sciences, Ottitsch and Krott [58] mention three central concepts for strategic decision-making about urban forests. “Polity” comprises the institutional dimension at formal levels relating to, for example, constitutions, laws, taxes, parliament, as well as at informal (e.g., traditions) levels. “Politics” then refers to the procedural dimension, looking at the dynamics of political processes, as, for example, concerning will formulation, interest mediation, bargaining, and communication processes. “Policy,” finally, refers to the substantial or normative dimension, which includes issues and objectives as well as outcomes of the political process. A policy document can be a typical outcome of the political process.

While politics and policy-making relate to the choice of and releasing of certain political decisions, planning encompasses a search for implementation options [34]. Planning has also been described as coordinating management in order to meet the objectives the community would like to see met [58].

In an urban forestry context, policy-making and policies define the strategic choices regarding urban forests, typically for a period of at least 10 years. These typically relate to which urban forest functions should be in focus, depending on societal demands. But urban forest policy-making is also about how to conserve and develop urban forests in highly dynamic, high-pressure urban settings. Conflicts are an integral part of urban forestry, and, not surprisingly, urban forest policy-making has thus been defined as a social bargaining process between stakeholders for the regulation of conflicts related to different interests concerning urban forest utilization and protection [58].

As urban forests comprise a form of urban land use, planning helps to relate urban forest to other types of land use, as well as helps guide choices within urban

forest areas. Planning can be seen as the intermediate level between decision-making (policy) and actual implementation of decisions. In urban forestry, the latter comprises, for example, the design and management activities described elsewhere in this chapter.

Urban forest policies and planning deal with a highly complex field, with a large variety of actors in play. Complexity also relates to the complexity of the urban forest itself. As is seen, it constitutes an overall green network that includes tree elements ranging from street trees to large woodland areas.

Urban Forest Policies and Plans Today

Research has shown that across the globe, urban forest policies are scarce when regarded in terms of formalized policies that specifically focus on a city’s “urban forest,” incorporating all tree-dominated vegetation [34, 60]. A growing number of cities in especially the industrial world have developed policies and strategies for (part of) their green spaces, as, for example, reflected by green structure plans. Green belt policies inspired by the London Green Belt have been in place in many parts of the world [2]. But a strategic perspective on what should happen with a city’s woodlands and trees is seldom present. Exceptions are North American and British cities with specific urban forest plans. The most typical level of plans, however, has been that of (park and forest) management plans which pay less attention to strategic, longer-term decisions, and focus more on short-term, operational maintenance tasks. There are exceptions, however, such as the strategic management guidelines developed for forests in and near major German cities.

As urban forest policies are largely absent at the local level, it is not surprising that they are even more scarce at the national level. Still, some national urban forest policy-making has occurred, with countries such as the USA and China as important examples. In the USA, a nationwide urban forestry program has helped raise political awareness about urban forestry, as well as aided implementation of urban forestry at state and local level (e.g., [13]). Moreover, the national program has been supported by substantial research funding. In China, urban forestry and urban greening policies have been developed by the national government. As part of the national programs, cities with ambitious and

successful greening programs being awarded awards and privileges [43]. Elsewhere, countries such as Malaysia and Singapore have developed ambitious national “greening” programs (e.g., [60, 77]) in which tree planting plays a major role.

In Europe, the UK has probably been the country with the most ambitious urban forestry policies. Urban and community forests were described as a priority and powerful tool in the England Forestry Strategy issued in 1998 [18]. In a national program called the Community Forest Program, woodland establishment and management near large agglomerations has been used to tackle significant social, economic, and environmental challenges [36]. The Community Forests were set up as partnerships between national and local public bodies, as well as a range of private actors. When focusing on the woodland element of urban forests, the last 2 decades have shown a boom in new policies that have incorporated the importance of urban forests and/or urban forest elements. Countries such as Belgium (Flanders), Denmark, Ireland, Netherlands, and Great Britain issued afforestation policies in which urban agglomerations have the highest priority. Woodland grant schemes thus favor urban settings. Social and environmental services such as providing opportunities for outdoor recreation and protection of drinking water for primarily urban populations have become prioritized in national forest policies.

At the international level, the Food and Agriculture Organization (FAO) of the United Nations started to actively engage with “urban and peri-urban forestry” during the late 1980s. From the late 1990s, efforts intensified (e.g., [38]). FAO, for example, assigned a series of urban forestry case studies in the developing world, demonstrating that urban forestry was not only a promising approach for western countries, but could help developing-country cities in dealing with environmental and economic challenges. FAO also assessed the status of urban forestry legislation and released a series of guidelines for good practice in, for example, urban forest policy-making. Urban forestry became integrated in the so-called Regional Outlook Studies for forestry, starting with the Forest Outlook Study for West and Central Asia [17].

Policy objectives for urban forestry have changed over time, in line with changing societal demands and focus on different urban forest benefits. During a large

part of history, for example, especially urban woodland was conserved and managed as provider of local timber, fuelwood, and animal fodder [36]. City forests such as the Sihlwald near Zurich, Switzerland, for example, were gradually “annexed” by the neighboring city eager to guarantee a steady supply of timber for construction. Urban woodland areas were also protected by kings and other rules as hunting domains. Throughout history hunting was not only a favorite pastime of those in power but also a means of obtaining status and prestige. Prestige was also an important reason for developing the first city parks and boulevards, inspired by the work of Haussmann in Paris.

Over time, urban forestry became targeted toward a wider range of users, with the benefits of green space, for example, for public health and worker well-being becoming recognized in the industrialized world. Recreation and other social benefits came in focus – where they still are today, as described elsewhere in this chapter. This can be illustrated by various studies of urban forest policies in Europe (e.g., [34, 58]). The social benefits of urban forests are prioritized in most cities, with environmental services such as protection of drinking water, improving the urban climate and nature protection ranking second. Economic benefits in terms of, for example, timber production is only of secondary importance in most western cities, although several cities in, for example, Central Europe do manage their forests for timber. Moreover, new focus on bioenergy as alternative to fossil fuels also makes wood harvesting more attractive once again [36]. In developing countries, urban forests often make important contributions to local livelihoods.

Although there seems to be agreement that urban forests primarily have amenity values, this does not mean that conflicts over the green spaces and their use do not occur. In fact, urban forestry policy-making is very much about managing conflicts, for example, between a range of different recreational uses. Relatively small areas of urban forest near to cater for the wishes of thousands and sometimes millions of people.

Conflicts are also mentioned as an important element of urban forestry policies and planning by Ottitsch and Krott [58] in their study of urban forestry policies and planning in 14 European cities. They describe a number of common elements apart from a joint focus on social and environmental benefits.

The authors stress the importance of municipal land property and regulative policy instruments – and the fact that current urban forest policy mostly deals with publicly owned land, even though a considerable part of a city's urban forest is in private ownership (e.g., private garden trees). In the case of private ownership, public authorities use legislation such as tree protection orders and financial instruments such as subsidies for private land owners offering recreational access to their lands.

Actors and Stakeholders

From a “polity” perspective, municipal authorities are the main decision-makers for urban forests. The strong role of municipal actors is partly due to what Ottitsch and Krott [58] call the traditional “property strategy” according to which city governments attempted to transfer green space within their boundaries or within their otherwise described spheres of interest to municipal property.

In spite of the dominant role of municipal governance, however, decision-making about urban forests is seldom a comprehensive process with one central actor. Typically, different components of the urban forest are the responsibility of different municipal departments. Comprehensive policies are difficult to develop, as city trees, parks, and woodland areas are often placed in units or divisions each with their own culture and way of doing things. As Liu [43] has shown for the Chinese city of Weihai, city park and forestry units have had difficulties in cooperating or at least followed their own agendas. A trend in many European countries has been, however, to integrate the different green space elements (e.g., parks, nature areas, and forests) in one division or department, something which could be expected to lead to more comprehensive urban forest policies and planning.

Other public actors also play a role in urban forestry decision-making, especially through state forest agencies that manage forest areas surrounding cities and towns. In some cases, state forests are even part of the city proper. Often these state forests are former royal hunting domains [34].

Private land owners do play a role in urban forestry, for example, through privately owned wooded estates. Moreover, a large part of a city's tree population is typically located in private gardens.

In addition, urban forestry policy-making and planning involve a wide range of other stakeholders. Obviously residents and users play an important role, as primary beneficiaries of urban forestry. Participation and communication regarding urban forest issues has steadily increased during recent decades. As Moll [52] states, “Where we find healthy city forests, we usually find strong citizen support. In the end, the support of city managers and government agencies comes back to citizens. They are the conscience of the community and the clients all political leaders work for. No other group feels the sense of place trees offer or organizes a force to change conditions more effectively than citizens. City forests decline when city foresters consider citizens a group to avoid, and the forest improve when lines of communication are opened.” Although public participation has been increasingly called for in urban forestry, it is not always a well-developed component of decision-making processes [58, 85]. Another important group of policy actors are nongovernment groups such as nature conservation and outdoor recreation associations.

Policy Integration and Governance

Urban forest policy-making and planning are changing in response to a series of challenges to urban forests and local authorities. Ottitsch and Krott [58] speak of a number of “tension lines” in urban forest policies. In many cases, as described, there are no comprehensive policies for urban forests, but rather a patchwork of segmented policies, different spheres of interests, and competition between different municipal bodies. The role of urban forestry as a public service is also under debate. Urban forests are seen as a public service and public good, which has led to focus on expensive property strategies. In a time of public budget cuts, alternative funding mechanisms and policy instruments need to be developed. Conflicts over urban forests (and their use) have intensified and urban demands are rapidly changing.

However, urban forestry also involves a range of promising opportunities. The role of trees and other vegetation in tackling major political issues such as climate change mitigation and adaptation, promoting human health and well-being, and enhancing city competitiveness in a globalized society is increasingly

recognized. Forests and trees can play a role in large, market-driven development projects, for instance for new sustainable urban development schemes. Through the strong associations people have with forests and trees, these can play an important role in symbolic communication. One example is representing a physical sign of how the local environment is improving [42]. This symbolic role has been intensively used in programs such as the English Community Forests, but also in the “rebranding” and greening programs of a city like Chicago. In a time of, on the one hand, individualization, and on the other hand migration, urban forests are among the final truly public areas where all of society’s segments and groups can meet each other. Thus urban forests can help strengthen social cohesion (e.g., [62]).

Developing the potential of urban forestry requires closer collaboration between stakeholders. Municipal authorities alone can no longer manage the task alone, hampered by government reform, lack of finances, and calls for greater involvement. New modes of governance are needed – with governance being defined as the process of making decisions that define expectations, grant authority, and verify performance [33]. In urban forestry, the traditional approach of “governance by government” needs to be (and is increasingly) replaced by modes of “governance with government,” as, for example, reflected in public–private partnership and public involvement processes.

New forms of governance for urban forests are being explored. In Finland and Sweden, for example, National Urban Parks have been established. The Stockholm National City Park, which includes the Norra Djurgården city forest, was the first of its kind. The park, established by parliamentary decision in 1994, is the result of new environmental and land use legislation and the national level. National urban parks are developed as a partnership between state and municipal authorities, as well as other stakeholders. The overall aim for the NCP is to act as a model for sustainable development, with focus on nature and ecology, recreation and culture. Swedish legislation stipulates that national urban parks should contain natural areas important for the preservation of urban biodiversity, cultural milieus – including buildings – important for an understanding of national history or that of the city itself, and parks and green areas of

architectural or esthetic significance [71]. Multi-stakeholder schemes for urban forests of national as well as local importance also include the Biosphere Reserves developed near cities from Vienna to Sao Paulo [36].

Urban forestry can “feed in” to emerging, integrative planning approaches such as “green infrastructure planning.” The term “green infrastructure” has become frequently used in countries such as the UK and the USA in reference to urban renaissance and green space regeneration. It can be defined as creating networks of multifunctional green spaces that are carefully planned to meet the environmental, social, and economic needs of a community [44]. All green spaces, both public and private, are considered. By using the term “infrastructure,” focus is on functionality and green spaces enter the urban planning discourse at an essential public service.

Design and Management of the Urban Forest

As reflected in the overall concept of urban forestry as described above, design and management of urban forests is an interdisciplinary and participatory field at the interface of, among others, landscape architecture, landscape planning, horticulture, arboriculture, forestry, ecology, and the social sciences. Whether urban forest elements are of natural, seminatural, or man-made origin, and whether new elements need to be added to the urban green structure or existing components require changes due to evolving demands, proper design and management are needed to ensure that woodlands, parks, and trees in streets and other urban spaces can provide the environmental, economic, and social functions society attributes to and demands from the urban forest resource. The disciplines, professions, and stakeholders involved vary depending on the location, setting, and social and cultural context, and not at least on the type of urban forest element concerned.

Across the world, urban forest design and management relate to very different circumstances. The climate, the ecosystem, the character of the people, the cultural history, city development, and the political settings over time varies between cities, between regions, between countries, and between the continents. Design and management aim to maintain and

incorporate the rich variety of climate, ecology, tradition, and cultural heritage contained in the existing resources, whilst at the same time meeting current demands and incorporate flexibility for changes in future requirements (population pressure and social values) and climate (climate change) [5].

Many of the most appreciated urban woodlands, parks, street tree plantations, and private gardens are characterized by old and graceful trees recognized for their cultural–historical and natural values. An inventory of the Belvoir Park Forest at the fringe of Belfast, Northern Ireland, for example, found 270 single stem trees with a diameter at breast height of 3 m or more [75]. Such trees act as landmarks and offer important habitats for biodiversity [31] and management often aim to maintain them for as long as they do not pose a security risk [28]. Design and management of urban forest therefore have to consider the full life cycle that trees have in nature, including developing stages of establishment, growth, early maturity, and degrading stages of late maturity and degradation, and today continuous design and redesign over time is perceived as a crucial part of sound, sustainable urban forest management [5, 25].

Whereas the urban forest traditionally has been designed and managed with the mature stage in mind only [24], the pace of urbanization and social changes leading to cities being transformed and renewed faster than trees can grow, has resulted in a changed professional mindset. Design is increasingly perceived as a continuing process rather than a one-off action. It is seen as a crucial part of sound, sustainable management [5, 25]. This has led to an upsurge in integrative and adaptive approaches in urban forest design and management that focus on developing qualities through all stages of development, from establishment and growth through maturity and to degradation (e.g., [10, 53]), while in streets, squares and other “gray spaces” the demand for “instant” values have implied that planting of large trees of 20–30 cm circumference or more, is becoming increasingly common practice [79].

Design of urban forests operates on different levels [5]. At the integrative and strategic level, design operates on the regional scale where urban green areas are linked to the landscape structure of the wider region, such as the “Finger Plan” of the greater Copenhagen area in

Denmark. At the city level, urban green structures are designed to give strength and depth to the urban forest by linking individual components and their form and functions. Frederik Law Olmsted’s design of the “Emerald Necklace” in Boston, Massachusetts, from the 1870s is one of the first and most known examples of design constituting a green infrastructure. At the site level, the form, structure, and content of individual green spaces are designed with linkages to local aspirations and use patterns, including the detailed layout and use of different vegetation elements, such as the number of plants involved, what types they are (trees, shrubs, herbaceous plants), the mix of species, how the plants are assembled in space, and their age or maturity.

The concept of urban forest management relates to management as operations and management in terms of organizations carrying out the management [25]. Municipal authorities are the main managers of urban forests but following the introduction of New Public Management doctrines a growing number of cities in especially the industrial world have outsourced management to private enterprises (e.g., [65]). Trees in private or commercial spaces such as gardens or business premises often represent a significant proportion of the tree population of a town or city. Although it is difficult to control how trees on private property are used and managed, urban foresters are increasingly becoming aware of the need of including them in the complete urban forest resource. During the last decades many cities throughout the world have developed and implemented orders and legislations that protect and prohibit removal or damage of trees in public and/or private areas by use of, for example, felling licenses, replacement planting regulations and enforcement systems such as penalties [64, 73].

Management operations address different levels. Gustavsson et al. [25] write “At the level of strategic management, the overall visions for management are developed. In line with policies and planning, objectives and targets are formulated, means are allocated and a time frame is set. Specific, well-defined tasks are defined and carried out in line with this at the level of operation management. While strategic management typically addresses a period of 10 years or more, operational management focuses on annual or biannual activities. As an intermediate level, tactical management brings the two together.”

Design and management of urban forests are directed by objectives and targets, economic and other frames, and not at least the type of green area concerned. The variety of urban situations that trees relate to has been grouped into three types [66]. The first type of location is trees in street, square, and other “gray spaces” with sealed surfaces. The second include parks and other green spaces such as yards, gardens, and commercial areas. The third is stands of trees which are referred to as “woodlands” or “woods” to distinguish between the wider urban forest resource and its components that have traditionally been defined as “forest.” The main design and management issues, concepts, and principles vary between the three types of locations.

Trees in Streets and Other Paved Areas

Row plantings of uniformly spaced trees are one of the earliest and simplest expressions of intentional and functional design. Over the ages, trees have been an integral part of the urban infrastructure and a mirror of the prosperity and achievements of society, such as the *Platanus* alley in the Agora of Athens, 500 BC. In Europe, the use of actual street trees became more widespread following Haussmann’s renewal of Paris during the 1850s and beyond. Here trees became substantial street fixtures perceived as providing important esthetic and climatic benefits [49], and today systematic tree planting in streets and boulevard is a powerful characteristic of the city.

Trees in streets and paved areas are an important component of the urban forest, as well as an important architectural element in the cityscape where they ameliorate esthetic, social, and microclimatic conditions [3]. The natural, graceful shapes of trees provide a transition between human size and the scale of buildings and streets, as well as an element of continuity and uniformity in heterogeneous districts of mixed uses, building styles and variable scales [5]. Trees can be designed to relate to built form in different ways by use of different architectural principles of scale, massing, enclosure, and perspective, by use of different design concepts (e.g., boulevard, avenue, alley, single line, block, grid, or groups design) as well as different functional characteristics related to tree from size, texture, color, and different degrees of light and shade

[3, 20]. In many cities, the “web” of trees along streets, squares, etc., weaves through the city and supports the overall sense of a citywide forest by creating ecological corridors and visual links between woodland, parks, and other open spaces. In the historical or commercial core of larger cities, streets and not in the least parking areas are often the only location available for trees. In countries with a warm climate, the shade provided by trees encourages life to take place in the street.

Streets and other paved urban areas with surface sealing are generally characterized by limited space for planting pits, alkaline and compaction soils resulting from construction work, utility trenching, and in the many parts of the world use of deicing salt. The complex stressed environment challenge tree growth [74] and integration of adequate spaces and growth substrates in the overall street design is increasingly called for [61, 76], as is the use of stress-tolerant tree species (e.g., [4, 79, 80]). In the Central European region, an important contribution is the Climate-Species-Matrix developed by Roloff and colleagues [67] to support choice of tree species for urban habitats considering climate change. The matrix contains 250 woody species used in Central European green spaces. Based on drought tolerance and winter robustness, each species is assessed and classified with regards to usability after predicted climate changes of increased average temperatures and more frequent heat waves and periods of drought.

Research has shown that a limited number of tree species dominate the tree stock in streets, but drawing on the severe consequences of Dutch elm disease a growing number of experts emphasize the use of a diversity of species as key component of design strategies aimed at increasing the resilience of the tree stock to pests and diseases [4, 51]. The most comprehensive recommendation of the latter concept is provided by Santamour [70], who suggests that no species should represent more than 10%, a genus not more than 20%, and a family not more than 30% of the population.

Besides selection of suitable tree species and proper design and planting, aftercare is essential to maintain healthy, esthetic, and safe trees in the harsh conditions of streets and other urban paved sites. Urban tree management techniques and activities are grouped under the field of arboriculture, that is, the cultivation and management of individual or small groups of

woody plants, primarily in urban areas. Arboriculture includes watering, fertilization, mulching, protection with (stakes, supports), and formative pruning of young plantings, as well as training, vitality, and safety inspections of mature trees to reduce hazard [9]. It is based on a detailed understanding of tree biology and responds to cultural practices of pruning and to wounds and environmental conditions. Arboriculture also plays an important role in management of trees in parks, gardens, and other urban green spaces, but here, management of other types of vegetation also comes in.

Parks and Other Green Spaces

Parks, gardens, and other urban spaces dominated by vegetation often contain a considerable number of trees, but actual forest stands are often limited or absent. Instead elements such as lawns, pastures, garden elements such amenities as seating and different types of infrastructure and/or buildings dominate [5]. Design and management aim to integrate elements into attractive areas that are enticing for people of all ages, genders, and backgrounds, and stimulate use on regular basis during all parts of the day, week, and year, where trees are used to provide ornamental values as well as mass and structure and define the character and scale of space.

Park design has evolved through a number of phases where different styles have reflected the social, cultural, and artistic movement of the times when they were created. Beautification has traditionally been one of the main arguments for greening cities with parks, gardens, and other types of green spaces [36, 82]. The park first became an integral part of city planning and development in England during the 1830s, and then subsequently spread throughout Europe, North America, and eventually the world. In the polluted and densely populated cities that followed the industrial revolution, parks were seen as a solution to both social and hygienic problems. Peoples could gather and socialize in the fresh air, which promoted public health [8]. Playgrounds, sports fields, and other facilities stimulation activity evolved as one of the most important features of public parks, and remain so today. Yet, as evidence has mounted about the multiple functions urban green spaces provide, their design and management has gradually shifted to also encompassing

environmental, conservation, and economic benefits with the aim to develop functioning ecosystems as well as settings for cultural and social activity [5, 25, 47].

Recognizing the interconnection of natural resources and human resources and design and management of parks and other green spaces have become essential for the concept of sustainable urban ecosystems. The political movements of sustainability and protection of biodiversity have caused a more restricted use of the richness of exotic species, which for centuries have been a predominant element in park design due to their ornamental values. The focus of decision makers, professionals, as well as residents on protecting local nature and cultural history have implied that the design and management of urban parks and woodlands increasingly are carried out as part of varying restoration programs which either concentrate on ecosystems functions or cultural history, or both (e.g., [22]). One example is the “Isar Plan,” where the strongly regulated Isar river course in the inner-city of Munich, Germany, has been redesigned as part of a program that combines flood protection with the restoration of a near-natural river landscape and the improvement of nature-oriented leisure and recreational use of the urban population in the river bank zone [16].

Trees in private or commercial spaces such as gardens or business premises often represent a significant proportion of the tree population of a town or city, and give people the chance to express themselves and to develop their own, personal relationships with trees and plants. Although it is difficult to control how trees on private property are used, urban foresters are increasingly becoming aware of the need of including them in the design and management of the complete pattern of urban woods, parks, street trees, and trees in private spaces [5]. In recognition of this, an increasing number of cities have developed tree protection orders that also include private trees [64, 73], as also described elsewhere in this chapter.

Woodlands

Urban woodlands are designed and managed for a variety of purposes. Originally, provision of timber and fuelwood was the main reason for cities and towns to own and manage forests, but this function is

presently only of secondary importance [36]. Recreation and nature protection typically are the main functionalities of urban woodlands today [5, 82]. Woodlands tend to be more multipurpose than parks and other urban forest elements, and their design and management need to accommodate more uses, some of which might conflict. Aspiration of woodland character differs greatly between users, for example, according to age, sex, profession, and cultural background, and it is important to consider the social status and recreational use of the individual woodland as well as the ecological status and potential for nature conservation, and to rank the use priorities when deciding on their design and management [28, 46].

Bell et al. [5] define woodland design as "...accomplished through the manipulation of the spatial pattern of different stands of trees, their species composition and vertical structure(...). These elements are manipulated through woodland management and the adoption of different silvicultural systems." However, the importance of urban woodlands has from a silvicultural perspective been widely overlooked and approaches that are suited to the capacities of woodlands and the realities in urban contexts are still under development. Much urban woodland has developed over time from commercial production plantations and mainly consists of even-aged monoculture types with interiors resembling the "pillared hall." However, canopy trees, undergrowth, and the main characteristics of the ground flora are all of importance for biodiversity and for the human experience of character, size, and atmosphere of forest interiors [25]. Municipal authorities throughout Europe and North America are increasingly applying a variety of management techniques to achieve different structures of vegetation and atmospheres within individual woodlands as well as across woodland lots. Form- and species-rich forest stand types associated with selection cutting methods are considered among the most valuable types to integrate high esthetic and biological qualities [25, 90], especially in smaller-scale woodland lots. Historical/cultural forest management regimes, such as coppice forests, forest pastures, and unmanaged forests are also increasingly applied to assist diversifying woodland interior rooms and habitats within and across urban woodlands for the benefit of people, plants, and wildlife [25, 54, 69]. Natural colonization has also been

identified as a management tool that can give strength to urban woodland design and that is readily transferable, although regional differences have to be taken into account [68].

Compared to "commercial" forests with a management regime aimed at wood production, and where the stand is the functional unit, the higher functional level – that is, the woodland landscape – is of equal importance when it comes to developing recreational and ecological functionalities in urban woodlands [25]. Hence, design and management of urban woodlands should not only give attention to stands, but also to the wooded landscape. Variation between wooded areas and open habitats and recreational areas is a common characteristic for many of the most appreciated and well-known urban woodland landscapes in Europe, such as Amsterdamse Bos, the Netherlands, Helsinki's Central Park, Finland, and Jægersborg Deer garden north of Copenhagen, Denmark. The mosaic of interiors under closed canopies, intimate glades, and canopy openings, semi-open savannah-like areas with scattered trees, open spaces of varying sizes and types and a diversity of edges along the boundary of forested and open parts are mutually supportive for multipurpose uses [5], while they also create habitat diversity, something which is of the utmost importance for biodiversity conservation.

Urban growth has in many cases led to woodland fragmentation, cutting woodland areas into more or less isolated patches of varying sizes, ranging from small copses and woodland remnants to large wooded landscapes. This fragmentation should be considered when designing and managing urban woodlands. While larger woodland landscapes are often the most used for recreational purposes, small wooded lots account for a large share of the woodland resource in many cities. In the case of Ljubljana, Slovenia, for example, Pirnat [63] identified as much as 606 woodland lots of a size of 0.5–10 ha. While woodlands in general are able to absorb many uses, small woods limit recreational as well as ecological opportunities whilst larger woods can absorb many uses and provide a wider range of ecosystem services [1, 5]. Zoning different functions and applying different silvicultural systems in order to separate and provide attractive and safe environments for various forms of recreation and

ecosystem functions within woodland lots and across the diversity of urban woodlands are frequently applied [19, 23, 69].

Urban Forest Information and Inventory

To support sound urban forestry programs, policies, planning design, and management, high quality information is needed. Schipperijn et al. [72] divide the wide range of information requirements into basic green space information, environmental and ecological information, and sociocultural information. The need for reliable information has led to the development of different methods, tools, and systems, such as remote sensing, tree inventory systems, biotope mapping, sociotope mapping, GIS systems, and decision support systems to help collect, compile, and use available information, many of which are described in Schipperijn et al. [72]. Following the introduction of New Public Management doctrines, an increasing number of cities in especially the industrial world have outsourced urban green space management or implemented internal division of purchasers and providers in order to reduce management costs and to gain more insight into the daily management routine tasks and costs (e.g., [65]). In this process, detailed inventories and monitoring have been carried out to establish “basic green space information” about the location of green areas, green elements, and street trees, characteristics and quantity of different elements (species, age, size, height, and so forth), as well as information on type and quantity of the different management activities in use. In comparison, inventory and monitoring providing “information about the environmental and ecological” conditions that influence green spaces, as well as the environmental benefits of green spaces are still under development and implementation, as are methods for collect and compile “sociocultural information” dealing with user preferences, cultural values, economic benefits, and information on health and other social services.

Future Directions

The world’s urban population will only increase in the coming years, which means that the focus on cities as sustainable human habitats must increase as well. There will be a growing need for research and policy

support in the field of sustainable urban development. What should cities of the future look like? What constitutes a “good city”? How much green space is necessary to sustain a healthy population? Each city has its own specific problems, mainly related to economic status and geographic location.

At present, research results indicate that the urban population in general would benefit from a well-developed green infrastructure, especially when the urban green space is well integrated in the overall city planning. It is important to realize that planting more trees or creating more green in urban areas is not enough. Studies indicate the importance of well-funded, long-term management and proper maintenance of the green infrastructure. It is also clear that public participation ensures the success of government investment and is an important tool in providing sustainable green space and in creating sustainable cities.

The increased interest in urban ecology and environmental problems will drive investment in parks and green areas over the coming decades. Most of these should be laid out in the larger and rapidly expanding cities of Asia, Africa, and Latin America to ensure an equitable distribution of green space and healthy, sustainable population growth. Similar expansion of protected green infrastructure was carried out in North America and Europe during the years after the Second World War and up to the 1970s. Planning has not kept up with growth in many cases, and the need exists to look at a mega-regional scale to identify remnant urban and peri-urban open space and critical lands for protection that will be sufficient to supply a steady source of ecosystem services to sustain long-term population growth. Over the coming years, the main challenge will be to ensure that the expansion of green infrastructure in the world’s urban areas is implemented within the framework of sustainable development, optimizing the ecosystem services provided by these natural resources with the use of appropriate technology, community participation, and without a heavy dependence on chemicals and irrigation. This will ensure that the rapidly expanding urban landscapes remain resilient and healthy habitat for a growing society.

Today’s focus on climate change will raise the demands for more compact cities. But a compact city with a lot of sealed surface and lack of green space is

very vulnerable for the effects of climate change. Therefore, the issue of urban density is a good example of a potential conflict between mitigation and adaptation concerns. If increasing density, in order to reduce energy by lowering travel demand and heating requirements, leads to the consumption of green space resources, its consequence will be the loss of a vital adaptation resource. The concept of the city of tomorrow is such an oxymoron. The challenge is how to combine the need for a compact city with people's need for green space close to where they live – The Green Compact City. *Inter alia* Scandinavian planning and building tradition contains several examples which could inspire. Better coordination between transport, land use, and open space planning; urban containment and preservation of the green infrastructure; the creation of new urban landscapes; and promotion of the urban–rural interface are basic principles for this. Urban forestry is one of the promising approaches to support this process.

Bibliography

- Alvey AA (2006) Promoting and preserving biodiversity in the urban forest. *Urban Forest Urban Greening* 5:195–201
- Amati M (ed) (2008) Urban green belts in the twenty-first century. Ashgate, Aldershot
- Arnold HF (1993) Trees in urban design, 2nd edn. Van Nostrand Reinhold, New York
- Bassuk N, Curtis DF, Marranca BZ, Neal B (2009) Recommended urban trees: site assessment and tree selection for stress tolerance. Urban Horticulture Institute, Cornell University, Ithaca, New York
- Bell S, Blom D, Rautamäki M, Castel-Branco C, Simson A, Olsen IA (2005) Design of urban forests. In: Konijnendijk CC, Nilsson K, Randrup TB, Schipperijn J (eds) Urban forests and trees. Springer, Berlin, pp 149–186
- Benavides Meza HM (1992) Current situation of the urban forests in Mexico City. *J Arboric* 18:33–36
- Bernatzky A (1994) Baumkunde und Baumpflege. Thalacker Verlag, Braunschweig
- Bucht E (2002) Traditions in urban park planning and management in Sweden and other European countries. In: Randrup TB, Konijnendijk CC, Christophersen T, Nilsson K (eds) Urban forests and trees, Proceedings No.1. COST Action E12. Office for Official Publications of the European Communities, Luxembourg, pp 215–226
- Dujesiefken D, Dreno C, Oven P, Stobbe H (2005) Arboricultural practices. In: Konijnendijk CC, Nilsson K, Randrup TB, Schipperijn J (eds) Urban forests and trees. Springer, Berlin, pp 419–441
- Dunnet N, Hitchmough J (eds) (2004) The dynamic landscape. Spon Press, London
- EEA (European Environment Agency) (1998) Europe's environment: the second assessment. Office for Official Publications of the European Communities, Luxembourg
- Ellefson PV (1992) Forest resources policy: process, participants, and programs. McGraw-Hill, New York etc
- Elmendorf WF, Cotrone VJ, Mullen JT (2003) Trends in urban forestry practices, programs, and sustainability: contrasting a Pennsylvania, US, study. *J Arboric* 29(4):237–246
- Escobedo F, Chacalo A (2008) Estimación preliminar de la descontaminación atmosférica por parte del arbolado urbano de la Ciudad de México (A preliminary estimate of air pollution removal by Mexico City's urban forest). *Interciencia* 33:29–33
- Escobedo FJ, Wagner J, Nowak DJ, de la Maza CL, Rodríguez M, Crane DE (2008) Analyzing the cost-effectiveness of Santiago Chile's policy of using urban forests to improve air quality. *J Environ Manage* 86:148–157
- European Urban Knowledge Network (2010) European Urban Knowledge Network Web site, <http://www.eukn.org>. Retrieved 3 March 2010
- FAO (2006) Urban and peri-urban forestry and greening in West and Central Asia: experiences, constraints and prospects. Compiled by Åkerlund U, Knuth L, Randrup TB, Schipperijn J. FAO, Rome
- Forestry Commission (1998) England forestry strategy. A new focus for England's woodlands – strategic priorities and programmes. Forestry Commission, Cambridge
- Forestry Commission (1992) Forest recreation guidelines. HMSO, London
- Forrest M (2006) Landscape trees and shrubs, use and management. CABI, Wallingford
- Gill S, Handley J, Ennos R, Pauleit S (2007) Adapting cities for climate change: The role of the green infrastructure. *Built Environ* 33(1):115–133
- Gobster PH (2007) Models for urban forest restoration: human and environmental values. In: Proceedings of the IUFRO conference on forest landscape restoration, Seoul, Korea, May 14–19, pp 10–13
- Gundersen VS, Frivold LH (2008) Public preferences for forest structures: A review of quantitative surveys from Finland, Norway and Sweden. *Urban Forest Urban Greening* 7:247–258
- Gustavsson R (2004) Exploring woodland design: designing with complexity and dynamics – woodland types, their dynamic architecture and establishment. In: Dunnet N, Hitchmough J (eds) The dynamic landscape. Spon Press, London, pp 246–293
- Gustavsson R, Hermy M, Konijnendijk CC, Steidle-Schwahn A (2005) Management of Urban Woodland and Parks – Searching for Creative and Sustainable Concepts. In: Konijnendijk CC, Nilsson K, Randrup TB, Schipperijn J (eds) Urban forests and trees. Springer, Berlin, pp 369–397
- Harris RW (1992) Arboriculture: integrated management of landscape trees, shrubs and vines, 2nd edn. Prentice Hall, Upper Saddle River
- Helms J (ed) (1998) Dictionary of forestry. Society of American Foresters, Bethesda

28. Hodge SJ (1995) Creating and managing woodlands around towns, vol 11, Forestry Commission Handbook. HMSO Publications Centre, London
29. Jensen FS (1995) Forest recreation. In: Hytönen M (ed) Multiple-use forestry in the Nordic countries. The Finnish Forest Research Institute, Helsinki, pp 245–278
30. Jensen MB, Persson B, Guldager S, Reeh U, Nilsson K (2000) Green structure and sustainability – developing a tool for local planning. *Landscape Urban Plann* 52:117–133
31. Jonsell M (2004) Old park trees: a highly desirable resource for both history and beetle diversity. *J Arboric* 30:238–244
32. Kaplan R, Kaplan S (1989) The experience of nature. Cambridge University Press, Cambridge
33. Kleinschmit D, Böcher M, Giesen L (2009) Discourse and expertise in forest and environmental governance – an overview. *For Policy Econ* 11(5–6):309–312
34. Konijnendijk CC (1999) Urban forestry in Europe: a comparative study of concepts, policies and planning for forest conservation, management and development in and around major European cities. Doctoral dissertation, Research Notes No. 90, Faculty of Forestry, University of Joensuu, Joensuu
35. Konijnendijk CC (2003) A decade of urban forestry in Europe. *For Policy Econ* 5(3):173–186
36. Konijnendijk CC (2008) *The Forest and the City: the cultural landscape of urban woodland*. Springer, Berlin
37. Konijnendijk CC, Randrup TB (2004) Urban forestry. In: Burley J, Evans J, Younquist JA (eds) *Encyclopedia of forest sciences*. Elsevier Science, London, pp 471–478
38. Konijnendijk CC, Sadio S, Randrup TB, Schipperijn J (2004) Urban and peri-urban forestry in a development context – Strategy and implementation. *J Arboric* 30(5):269–276
39. Konijnendijk CC, Nilsson K, Randrup TB, Schipperijn J (2005) *Urban forests and trees*. Springer, Berlin, Heidelberg, New York
40. Konijnendijk CC, Ricard RM, Kenney A, Randrup TB (2006) Defining urban forestry – A comparative perspective of North America and Europe. *Urban Forest Urban Greening* 4(3–4):93–103
41. Kuchelmeister G (1999) Urbanization in developing countries – time for action for national forest programs and international development cooperation for the urban millennium. Forest policy research forum: the role of national forest programs to ensure sustainable forest management, 14–17 June 1999, Joensuu, Finland
42. Lawrence HW (2006) *City trees: a historical geography from the Renaissance through the nineteenth century*. University of Virginia Press, Charlottesville & London
43. Liu L (2009) Status and prospects for urban green structure planning in China – Weihai city as a case study. *Forest and landscape research* 41/2008. Danish Centre for Forest, Landscape and Planning, University of Copenhagen, Frederiksberg
44. Longhurst J (2007) Chain reaction. *Green Places* 2007(36):20–22
45. Louv R (2008) *Last child in the woods*. Algonquin, Chapel Hills
46. Lucas OWR (1991) *Design of forest landscapes*. Oxford University Press, New York
47. McPherson EG (2000) Urban forestry issues in North America and their global linkages. Presentation for the 20th Session of the North American Forestry Commission Food and Agriculture Organization of the United Nations, St. Andrews, New Brunswick, Canada. 12–16 June 2000
48. McPherson EG, Rowntree RA (1993) Energy conservation potential of urban tree planting. *J Arboric* 19(6):321–331
49. Meyer FH (1978) *Bäume in der Stadt*. Verlag Eugen Ulmer, Stuttgart
50. Miller RW (1997) *Urban forestry: planning and managing urban greenspaces*, 2nd edn. Prentice Hall, Englewood Cliffs
51. Miller RH, Miller RW (1991) Planting survival of selected street tree taxa. *J Arboric* 17:185–191
52. Moll G (1989) In search of an ecological urban landscape. In: Moll G, Ebenreck S (eds) *Shading our cities, a resource guide for urban and community forests*. American Forestry Organisation, Island Press, Washington
53. Nielsen AB, Jensen RB (2007) Some visual aspects of planting design and silviculture across contemporary forest management paradigms – perspectives for urban afforestation. *Urban Forest Urban Greening* 6:143–158
54. Nielsen AB, Möller FG (2008) Is coppice a potential for urban forestry? – The social perspective. *Urban Forest Urban Greening* 7:129–138
55. Nilsson K, Randrup TB, Wandall B, (2001) Trees in the urban environment. In: Evans J (ed) *The forest handbook*, vols. 1 and 2. Blackwell Science, Oxford, pp 347–361 and 260–271
56. Nilsson K, Baines C, Konijnendijk CC (2007) Health and the natural outdoors, COST Strategic Workshop in Larnaca, Cyprus, Final report. COST, Brussels
57. Nowak DJ, Crane DE, Stevens JC, Ibarra M (2002) Brooklyn's urban forest. United States Department of Agriculture, Forest Service, General Technical Report NE-290, Radnor
58. Ottitsch A, Krott M (2005) Urban forest policy and planning. In: Konijnendijk CC, Nilsson K, Randrup TB, Schipperijn J (eds) *Urban forests and trees*. Springer, Berlin, pp 117–148
59. Owen J (1992) *The ecology of a garden*. Cambridge University Press, Cambridge
60. Palijon A (2004) Urban forestry in Asia: state of the art. In: Konijnendijk CC, Schipperijn J, Hoyer KK (eds) *Forestry serving urbanised societies. Selected papers from the conference held in Copenhagen, Denmark, 27 to 30 August 2002*, vol 14, IUFRO World Series. IUFRO, Vienna, pp 25–40
61. Pauleit S (2003) Urban street tree plantings: Identifying the key requirements. *Proc Inst Civ Eng Munic Eng* 156:43–50
62. Peters K, Elands B, Buijs A (2010) Social interactions in parks: Stimulating social cohesion? *Urban Forest Urban Greening* 9(2):93–100
63. Pirnat J (2005) Multi-functionality in urban forestry – a dream or a task? In: Konijnendijk CC, Schipperijn J, Nilsson K (eds) *Urban forests and trees*, Proceedings No. 2. COST Action E12. Office for Official Publications of the European Communities, Luxembourg, pp 101–118

64. Profous GV, Loeb RE (1990) The legal protection of urban trees: a comparative world survey. *J Environ Law* 2:179–193
65. Randrup TB, Persson B (2008) Public green spaces in the Nordic countries: Development of a new strategic management regime. *Urban Forest Urban Greening* 8:31–40
66. Randrup TB, Konijnendijk C, Kaennel Dobbartin M, Prüller R (2005) The concept of urban forestry in Europe. In: Konijnendijk CC, Nilsson K, Randrup TB, Schipperijn J (eds) *Urban forests and trees*. Springer, Berlin, pp 9–20
67. Roloff A, Korn S, Gillner S (2009) The Climate-Species-Matrix to select tree species for urban habitats considering climate change. *Urban Forest Urban Greening* 8:295–308
68. Ryan J, Simson AJ (2002) Identifying good practice in the design of urban woodlands. *Arboricultural J* 26:309–331
69. Rydberg J, Falck J (2000) Urban forestry in Sweden from a silvicultural perspective: a review. *Landscape Urban Plann* 47:1–18
70. Santamour FS Jr (1990) Trees for urban planting: Diversity, uniformity and common sense. In: *Proceedings of the 7th conference of the metropolitan tree improvement alliance*, Lisle, pp 57–65
71. Schantz P (2006) The formation of National Urban Parks: a Nordic contribution to sustainable development? In: Clark P (ed) *The European city and green space*: London, Stockholm, Helsinki and St Petersburg, 1850–2000, *Historical Urban Studies*. Ashgate, Hants, pp 159–174
72. Schipperijn J, Pillmann W, Tyrväinen L, Mäkinen K, O'Sullivan R (2005) Information for urban forest planning and management. In: Konijnendijk CC, Nilsson K, Randrup TB, Schipperijn J (eds) *Urban forests and trees*. Springer, Berlin etc, pp 399–417
73. Schmied A, Pillmann W (2003) Tree protection legislation in European cities. *Urban Forest Urban Greening* 2:115–124
74. Sieghardt M, Mursch-Radlgruber E, Paoletti E, Couenberg E, Dimitrakopolus A, Rego F, Hatzistathis A, Randrup TB (2005) The abiotic urban environment: impact of urban growing conditions on urban vegetation. In: Konijnendijk CC, Nilsson K, Randrup TB, Schipperijn J (eds) *Urban forests and trees*. Springer, Berlin etc, pp 281–323
75. Simon B (2005) The history of Belvoir Park. In: Simon B (ed) *A treasured landscape – the heritage of Belvoir Park*. The Forest of Belfast, Belfast, pp 16–42
76. Simons K, Johnson GK (2008) The road to a thoughtful street tree master plan – a practical guide to systematic planning and design. http://www.forestry.umn.edu/extension/urban_com/street%20Tree%20manual.REVISED_2008.pdf
77. Sreetheran M, Philip E, Adnan M, Siti Zakiah M (2006) A historical perspective of urban tree planting in Malaysia. *Unasylva* 57(223):28–33
78. Stanners D, Bourdeau P (eds) (1995) *Europe's environment. The Dobris assessment*. Europe's Environment Agency, Copenhagen
79. Sæbø A, Borzan Z, Ducatillon C, Hatzistathis A, Lagerström T, Supuka J, Garcia-Valdecantos JL, Rego F, Slycken JV (2005) Selection of plant materials for street trees, park trees and urban woodland. In: Konijnendijk CC, Nilsson K, Randrup TB, Schipperijn J (eds) *Urban forests and trees*. Springer, Berlin, pp 257–280
80. Tello ML, Tomalak M, Siwecki R, Motta E, Mateo-Sagasta E (2005) Biotic urban growing conditions – threats, pests and diseases. In: Konijnendijk CC, Nilsson K, Randrup TB, Schipperijn J (eds) *Urban forests and trees*. Springer, Berlin, pp 325–365
81. Tyrväinen L (1999) Monetary valuation of urban forest amenities in Finland. Doctoral dissertation. Research Paper No. 739. Finnish Forest Research Institute, Vantaa, Finland
82. Tyrväinen L, Pauleit S, Seeland K, de Vries S (2005) Benefits and uses of urban forests and trees. In: Konijnendijk CC, Nilsson K, Randrup TB, Schipperijn J (eds) *Urban forests and trees*. Springer, Berlin, pp 81–114
83. Ulrich RS (1984) View through a window may influence recovery from surgery. *Science* 224:420–421
84. Ulrich RS, Simons RF, Losito BD, Fiorito E, Miles MA, Zelson M (1991) Stress recovery during exposure to natural and urban environments. *J Environ Psychol* 11:201–230
85. Van Herzele A, Collins K, Tyrväinen L (2005) Involving people in urban forestry – a discussion of participatory practices throughout Europe. In: Konijnendijk CC, Nilsson K, Randrup TB, Schipperijn J (eds) *Urban forests and trees – a reference book*. Springer, Berlin, pp 207–228
86. Williams NSG, Schwartz MW, Vesik PA, McCarthy MA, Hahs AK, Clemants SE, Corlett RT, Duncan RP, Norton BA, Thompson K, McDonnell MJ (2008) A conceptual framework for predicting the effects of urban environments on floras. *J Ecol* 97:4–9
87. World Resources Institute (2001) *Earthtrends: the environmental information portal*. World Resources Institute, Washington DC
88. Yang J, McBride J, Zhou JX, Sun ZY (2005) The urban forest in Beijing and its role in air pollution reduction. *Urban Forest Urban Greening* 3:65–78
89. Zipperer WC, Wu J, Pouyat RV, Pickett S (2000) The application of ecological principles to urban and urbanizing landscapes. *Ecol Appl* 10:685–688
90. Larsen JB (2000) From plantation management to nature-based silviculture – a Danish perspective, Third International Congress on Sustainability in Time and Space. *Pro Silva Europe* 187

Urban Redevelopment and Quality of Open Spaces

JULIANE MATHEY¹, DIETER RINK²

¹Leibniz Institute of Ecological and Regional Development (IOER), Dresden, Germany

²Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany

Article Outline

Glossary

Definition of the Subject

Introduction

Dimensions of Shrinkage and Vacant Land in East Germany

The Joint Program of the German Federal and State Governments: "Urban Redevelopment East"

The Demands Made of the New Open Spaces

Strategies and Tools for Open Space Development

Conclusion: New Open Space Qualities Through Urban Redevelopment?

Bibliography

Glossary

Brownfield Brownfield is an unused urban site of derelict or underused land, which require intervention to bring it back to beneficial use and which may have real or perceived contamination problems (see also: <http://www.cabernet.org.uk>).

Perforated city The concept is used in the sense of a city whose physical fabric contains a scattering of vacant sites and which is undergoing progressive redistribution.

Renaturation The concept is employed here not in the narrower nature conservation sense but as used by the Federal Ministry for Building, Transport and Housing, and the Federal Office for Building and Spatial Planning to refer to urban redevelopment projects that are basically concerned with the temporary or permanent downzoning of building land into green and open spaces [36].

Shrinking city Shrinking City can be defined as an urban area that has faced a population loss in large parts of it for more than 2 years and is undergoing economic transformations with some symptoms of a structural crisis.

Urban Redevelopment Program "Stadtumbau Ost" ("Urban Redevelopment East") Federal funding scheme in Germany which has the aim to demolish a certain number of flats in order to regulate the housing market.

Definition of the Subject

Urban shrinkage has become a new normality for a growing number of European cities and urban regions. It is a result of different but strongly interconnected processes: uneven economic development, demographic change, shifts in land use and urban form, as

well as housing preferences and lifestyles. Shrinking urban regions develop their own patterns of development, and form distinctive dynamics that differ from those of their growing counterparts. Shrinkage, however, produces brownfield sites or open spaces of various sizes practically everywhere in the urban fabric. Many local authorities engage in scattered downsizing within existing settlements. But this type of urban redevelopment reaches its limits when "holes" steadily multiply and the physical urban fabric disintegrates. Redevelopment and downsizing raise the question of how the design of open and green space can provide a new quality of urban development. The major issue for the future is how public funding can best be deployed in handling vast areas of previously developed land within increasingly perforated cities. This paper looks systematically at the solutions already available and what contribution they can make to improving the ecological quality and the quality of life in cities.

Introduction

The current discussion on shrinking cities stresses the long-term nature and irreversibility of shrinkage, necessitating the abandonment of long-established planning principles. Furthermore, note is repeatedly taken of the fragmentation and perforation of urban areas, caused not least by parallel shrinkage and growth. However, the discussion has focused primarily on housing and to some extent on infrastructural problems. Abandoned industrial land and new brownfield sites in residential areas, in contrast, tend to be given only cursory attention – they are literally a "marginal problem."

Shrinking cities are not a new phenomenon, but in the past they have been regarded only as isolated or regionally limited exceptions [1]. During the last decades shrinking cities have become a distinct urban process in particular in Europe [2]. While a considerable share of "urban Europe" has been growing during the last decades and continues to do so, other urban regions have been suffering from deindustrialization, population loss and decay for decades. Nearly one third of all of Europe's cities with more than 200,000 inhabitants have at least undergone one decade of population decline [3] in the last 45 years. Whereas some of these

cities have been losing population ever since the 1960s, the majority of cases started declining only in the 1980s and 1990s. While population losses and economic decline can also be found in numerous Western and Southern European cities, they are especially pronounced in Central Eastern and Eastern Europe [4]. Here, declining urban population numbers are rather the normal case than the exception [3]. It is a result of different but strongly interconnected processes, that is, uneven economic development, demographic change, as well as shifts [5] in land and resource use, urban form, housing preferences, and lifestyles. As these drivers of shrinkage will increasingly impact on urban and regional development in the future, shrinkage clearly represents a key regional development challenge in Europe. Recent statistics show that processes of shrinkage will even gain more importance over the next decades: According to estimations by the UN, the population of western, northern, and – most of all – southern Europe will decrease. Current calculations of the UN predict a decline in Europe's population from 729 million in the year 2000 to 628 million people in the year 2050.

Shrinkage has thus not only become a new normality for a growing number of European cities but a serious challenge for urban planning. In many cases, shrinkage is caused by the sudden vacating of waste sites or a systematic retreat from extensive development in the course of redevelopment. The question of the quality and use capability of the (new) open spaces therefore takes on new dimensions: How can or should previously developed land be integrated into the urban fabric? What contribution can it make to local residential amenity? What interim and after-uses can be envisaged?

This paper looks at the dimensions of shrinkage and vacancies in East Germany, where this process is in particular pronounced due to the interaction of outmigration, demographic change, and suburbanization. As a reaction to the shrinkage processes in East Germany, the federal government established the "Urban Redevelopment East" program, which seeks to create solutions to this development. It also investigates how open spaces in shrinking cities are treated, what solutions have hitherto been adopted, and what approaches are desirable from an ecological, social, and/or economic point of view. Examples are presented

from various cities in East Germany. The idea of the paper is that these solutions are applicable as well in other European cities or worldwide.

Dimensions of Shrinkage and Vacant Land in East Germany

In East Germany, 90% of all towns and cities with over 20,000 inhabitants have suffered a population decline ([6]; [7], p. 62). Shrinkage in such dimensions and at such speed is a new phenomenon in Europe [8].

It follows on from a process that began in GDR times. Between 1949 and 1989, the German Democratic Republic lost about two million inhabitants. Local authorities also lost population owing to government-planned concentration processes, so that there were old gaps in the fabric of many towns and cities, often in a derelict state. At present, three main processes are causing shrinkage in East Germany: (1) the deindustrialization that set in with regime change, (2) the suburbanization or disurbanization of the 1990s, and (3) demographic decline.

1. Deindustrialisation set in immediately after the change in regime, affecting practically all sectors of the industry and with dire consequences for the economic basis of East German towns and cities. Seventy percent of all job losses were due to the collapse of the industry ([9], p. 4). Thus many industrial and commercial sites fell vacant, especially in the late nineteenth-century urban extensions. An estimated 3–4% or more of land in urban areas, and up to 5% in industrial cities lies vacant (own estimates on the basis of data on Leipzig), and will probably remain unused in the future. Completely new dimensions in brownfield land in cities are being dealt with here – a process that will continue with the advance of urban redevelopment.
2. From the early 1990s to the mid-decade, the rapid expansion of the housing supply topped the list of local government priorities. In view of the housing shortage in East Germany and positive population forecasts, both the rehabilitation of old city centers and the construction of new housing were pushed. Housing was built everywhere with the aid of tax relief, write-offs, and government subsidies. Between 1991 and 1999 alone, 773,368 new dwellings were built [10], almost exclusively newly

constructed on greenfield sites. This led to parallel shrinkage and growth, stepped up land take, and, finally, a government subsidized housing surplus that now – again with government aid – has to be reduced.

3. Demographic decline in East Germany has been caused by outmigration and natural population development. Between 1989 and 2005, some 1.4 million people moved from East Germany to West Germany; if about 500,000 foreigners had not arrived, the migration loss would have been even greater [11]. In addition, there was a natural decline in the population of some 600,000 owing to the massive fall in the birth rate from 1990 ([11], p. 12). Over this period, the total population of East Germany fell by 11.7%. Further decline can be expected owing to demographic change. If the trend in East Germany persists, the population will have halved by 2050. More and more towns and cities, not only in East Germany but also in the western states, are expected to experience stagnation or shrinkage [12].

Owing to demographic developments and outmigration, as well as developments on the property and housing markets, some one million dwellings had fallen vacant by the end of the 1990s. But these vacancies were only one element in the shrinkage problem, which must be understood as a many-faceted upheaval involving economic, social, and cultural factors ([7], p. 58ff.). The many dimensions of this upheaval are beginning to emerge in outline. East Germany has embarked on a specific path in the development of settlement structures which finds little to compare with in West Germany. Some authors have described it as a “disurbanization process” (intra-regional deconcentration of the population in an overall context of demographic decline) [13]. Since shrinkage can be expected to continue in the long term, “reasonable strategies for a retreat from extensive development” are called for [14]. Cities already have to cope with the loss of functions.

The Joint Program of the German Federal and State Governments: “Urban Redevelopment East”

In response to housing vacancies, the federal and state governments set up an expert commission in the late

1990s on “Housing Structural Change in the New States,” on whose recommendation the federal government launched the “Urban Redevelopment East” program in 2002 [15]. The scheme was endowed with some €2.7 billion. By 2010, about 350,000 of the circa one million surplus dwellings were to be demolished. The demolition program is currently being implemented almost throughout East Germany. About 350 larger local authorities are now integrated in the program covering about 650 development areas, approximately three quarters of all East German towns and cities with populations of over 10,000 [6]. Most communities receiving support have a vacancy rate of about 10–15%, some even 20%. So far about 220,000 dwellings have been demolished (status end of 2009), reducing the vacancy rate to some 13%. Meanwhile, the demolition program is no longer limited to cautious downsizing: in many cities entire neighborhoods are being razed ([6], p. 11). Redevelopment has so far focused on large housing estates on the urban fringe. Some 40% of development areas are characterized by prefabricated high-rise housing from GDR times. Only one quarter have older housing stock dating from before 1948.

“Urban Redevelopment East” is clearly conceived as a consolidation program for the local housing market, but also pursues ambitious urban development objectives. Support is mainly provided for the demolition of housing, but the remaining stock, infrastructure, and the residential environs are to be upgraded, which includes the provision of new, attractive green structures. The aim is an integrated procedure, linking housing industry problems with urban development aspects. With an eye to ensuring the sustainability of the city as a whole, planned demolitions are to be embedded in urban structures and the neighborhoods involved are to be upgraded at the same time [16]. To ensure that downsizing, redevelopment, and upgrading works are coordinated and that demolition is adapted to the given urban structure, funding is made contingent on the elaboration of integrated urban development plans. Their purpose is “to coordinate the individual measures of urban redevelopment and combine them into a sustainable, meaningful whole” [17] and hence to improve urban amenity. The “Urban Redevelopment East” program thus sets high standards – also for the creation and design of

open spaces – which, however, both funding policy and local sectoral authorities have since substantially lowered. The thrust is now clearly toward rapid consolidation of the housing market. Funding for the after-use or conversion of cleared demolition sites are now scarcely available ([16], p. 758); in only one third of cases are upgrading measures undertaken, mostly planting. It is increasingly apparent that downsizing on the “onion-ring principle” from the periphery inward, as preferred by most local authorities, faces considerable difficulties in implementation. Urban redevelopment measures are now almost impossible to embed in an urban development concept, and in many cases they lead to the perforation of the urban structure. As a result, many isolated open spaces and vacant sites of varying sizes are created in the urban fabric. Perhaps for this reason every second local authority regards coping with previously developed land in central and significant locations as a serious problem [18].

In sum, the strategy of controlled shrinkage adopted by many cities has already come up against its limits, since it takes no account of these new needs but continues to be pursued for lack of feasible alternatives. Coping with existing and new brownfield land cleared in the course of urban redevelopment is proving to be a growing problem for which, as it has been seen, the program no longer provides anything like adequate funding.

The Demands Made of the New Open Spaces

The permanent rezoning of residential and industrial land, as well as infrastructure land in such dimensions, is a new urban development task. Initially, traditional demands are made of the new open spaces; they are to serve recreational purposes, for instance, in the form of small parks, green spaces, or playgrounds, and generally contribute to improving residential amenity. Ecological demands are formulated, for example, that habitat networks or green corridors should be created, thus contributing to nature conservation in the city.

Furthermore, new demands are advanced, notably that the new open spaces should help enhance the image of the locality and contribute to the positive presentation of the shrinkage process and its consequences. What is accordingly envisioned is not isolated,

planted plots: notions of a “new urban landscape” predominate in which the new open space elements are interlinked ([19], pp 645ff.). The hope is that this will not only improve residential amenity but encourage demand and investment, thus upgrading entire neighborhoods. The new open spaces are hence expected to help stabilize residential areas and counteract vacancies, thus maintaining urban and socio-spatial structures. Open space systems are to provide a solid framework for the maintenance of urban structural contexts ([20], p. 737ff.), help integrate peripheral settlements into the surrounding landscape, and to improve and create recreational and locational amenities in inner cities ([6], p. 23; [21]). One particular goal is to establish green corridors that serve both recreational and nature conservation purposes.

The urban redevelopment program treats the creation of green spaces primarily as an element in design upgrading; seldom are ecological or nature conservation goals in focus. Apart from municipal demands for improved residential amenity and better quality of life for local residents, the call for active participation by the public plays a steadily growing role. Top-down urban development is longer judged sufficient: the committed involvement of residents and private actors is also needed. This development also means that more is expected from the open spaces cleared during redevelopment.

In the course of urban redevelopment, neighborhoods, cities, and entire regions set their hopes in the new open spaces as “identity-forming elements” for “when buildings are demolished and entire neighbourhoods downsized, the only remaining key factor in urban development is open space” [22]. What is more, the uses to which open spaces are put change with population decline, aging, diversification, and individualization. The focus shifts increasingly to specific user groups: the elderly, youth, singles, families, women and men, Germans, and immigrants ([22], p. 7). The question is whether the new, differentiated demands can really be met by traditional amenities. They are replaced by uses not typical of the city, with open spaces being understood as creative spaces. They include experience, art, and nature areas such as wild gardens, farmland, event venues, etc. The new open spaces are particularly suitable for social intercourse, offering opportunities for a highly varied urban stage

set by the population. Security in such areas plays a major role, and problems can be caused by informal use, for example, as rubbish tips, dog runs, drug trafficking locales, or camp sites for the homeless.

Strategies and Tools for Open Space Development

The current after-uses for previously developed sites range from (complete) clearance or clearance with simple planting or afforestation without precisely defined use requirements, to the layout of permanent green areas (parks, green belts, sport grounds, playgrounds, and other amenities). The *renaturation* (see our definition of the term within the glossary) of such areas is increasingly envisaged for the urban fringe. For some years now, new strategies have been tested, especially in East German cities, experimenting with interim uses in the form of temporary greening (recreation, events, nature conservation), but also with unconventional solutions for permanent green spaces, such as the layout of compensatory areas, downzoning to farmland and urban forest, and urban wilderness. Support is given for the renaturation of land no longer used for building purposes by both the Federal Spatial Planning Act [23] and the Federal Nature Conservation Act [24]. Section 2 (2) (8) of the Federal Spatial Planning Act, listing the principles of spatial planning, lays down that “if land is no longer used on a permanent basis, the productivity of the soil shall be maintained or restored.” Section 2 (1) (11) of the Federal Nature Conservation Act prescribes that “Sealed surfaces which are no longer required shall be renatured or, where de-sealing is not possible or excessively expensive, they shall be left to natural development.” In addition, the “National Strategy on Biological Diversity” adopted by the federal government on November 7, 2007, explicitly calls for more greenery in cities [25].

The examples of current and new subsequent uses of urban brownfield sites and their urban development, aesthetic, and ecological value are discussed below.

(Complete) Clearance of Redundant Sites

The consequences of the abandonment of large industrial areas, especially from the early to mid-1990s, have still not always been satisfactorily overcome. In many cases it was sought to revitalize areas through the

massively subsidized demolition of buildings and complete clearance of sites, which preparatory land-use plans continued to register as industrial land. Where land could not be successfully marketed or only partially put to new use, unplanned reintegration of the area into the surrounding countryside was often the result, especially on the urban fringe. An undisturbed succession of vegetation thus took over large areas. Although interesting from an ecological point of view, this development met with a very mixed reception among nature conservationists owing the preponderance of ubiquitous species. Aesthetic aspects also gave cause for criticism, and acceptance among the local urban population was often lacking. Such derelict areas provoked anxiety, recalling the decline of entire industries and the associated loss of jobs ([26], p. 79) (Fig. 1).

Simple Clearance

Many downsizing projects funded by the “Urban Redevelopment East” program have already proved problematic, since they leave behind them nothing but simple, often dreary and derelict open spaces for whose maintenance no one accepts responsibility. Many local authorities have demolished or downsized single buildings within existing settlements (prefabricated housing estates, *Gründerzeit* neighborhoods), where, although the buildings have been removed, the technical infrastructure is left in place and sealed and streets retained. Vacant land is usually planted with lawn, and sometimes with a few shrubs or trees. As a rule, these areas are still zoned in the preparatory land-use plan as residential. In some redeveloped areas, especially prefabricated housing estates, vast land-use gaps have now opened up in the once closed structure of the city. Attempts to create extensive green corridors are often in vain; clearance sites tend to be scattered throughout the community. Their ecological value is thus low and their recreation value limited (Figs. 2 and 3).

Development of Permanent Green Spaces

A number of brownfield sites have, however, also been permanently converted into green spaces. Larger brownfield or clearance sites are linked up to form green corridors in the hope of improving social and ecological qualities, and, particularly on the urban



Urban Redevelopment and Quality of Open Spaces. Figure 1

Abandoned land with undisturbed succession on a former industrial zone along the River Elbe in Heidenau (Germany) (Photo: M. Arendt)



Urban Redevelopment and Quality of Open Spaces. Figure 2

Demolition of a prefabricated housing estate in Dresden-Prohlis (Germany) (Photo: R. Bendner)

outskirts, establishing a transition to the landscape. It is hoped that improving the quality of life and enhancing the image of the locality will help prevent urban sprawl [27]. One successful example of this strategy is the

Leipzig-Grünau green corridor (planning and implementation 1993–1996), where the abandoned route of a planned road bordering on rural areas offered favorable conditions for creating a green belt. After



Urban Redevelopment and Quality of Open Spaces.

Figure 3

Simply green former housing area with lawn, few shrubs, and trees in Dresden-Prohlis (Germany) (Photo: R. Bendner)

demolition of the prefabricated housing, new recreational amenities and green spaces with native trees and shrubs were created in the area, which had been short on greenery and open spaces. Local residents and institutions were involved in planning and design [28]. Such green spaces are generally of great value for recreational uses and, if well designed, can contribute to the ecological upgrading of the neighborhood or of the entire urban territory. Sustained conservation and maintenance can prove problematic, and more and more reliance is put on participation by local residents (Figs. 4 and 5).

Interim Use as Temporary Green Space

Ownership structures can make it difficult for municipal programs to reactivate smaller gap sites scattered throughout the city. Private property owners are wary

of the legal repercussions of “official,” even temporary, use of their plots and fear for the value of their building land. In order to induce landowners to tackle their vacancy problems, the “authorization agreement” has been developed [29, 30], a tool that provides for the meaningful, temporary public use of private brown-field sites without affecting existing building rights. A contract is concluded between the municipality and the property owner under which the owner makes unused sites available for concrete public uses (simple green areas, playgrounds, neighborhood gardens). In compensation, the municipality offers to pay for design and layout and limited maintenance and provides relief from property tax. Owners have accordingly been more willing to make land temporarily available for a wide range of uses, providing more opportunities for individually designed life spaces [29, 30]. Leipzig has pioneered the temporary use of open spaces under authorization agreements. The municipality has pursued a central objective: “more greenery, less density.” The “Guidelines for Urban Renewal” [31] calls for more and better green areas and open spaces. Under authorization agreements, Leipzig has so far managed to develop some 140,000 m² of new green areas and open spaces ([29], p. 46), covering a wide range of interim uses. Most temporary green space uses focus on recreation. Depending on how these areas are designed and used, they can also be of incidental and targeted value in attaining urban-ecological goals and offer opportunities for nature conservation in the city (Fig. 6).

Use as Compensatory Areas

The use of inner-city brownfield sites for ecological compensation purposes gives local authorities a valuable opportunity to assert the interests of nature conservation in urban life spaces. For they provide debt-strapped municipalities with a possibility to use the duty imposed by Section 1a (3) of the Federal Building Code to mitigate the impact of development on nature specifically for urban-ecological measures, hence improving quality of life for residents [32]. However, difficult ownership structures and often high preliminary financing costs mean that hardly any suitable sites are available. Ecological upgrading mostly involves de-sealing and planting works. However, the extent to which nature conservation areas can be used



Urban Redevelopment and Quality of Open Spaces. Figure 4

New designed green spaces in Leipzig-Grünau (Photo: Municipality of Leipzig)



Urban Redevelopment and Quality of Open Spaces. Figure 5

Public allotments in the Lehne-Voigt-Park in Leipzig (Photo: J. Mathey)

by the urban population is problematic. In order to steer impact mitigation compensatory measures into the inner cities, a number of municipalities have formed land pools that include previously developed

land. In Leipzig, for example, the city council decided that 50% of compensatory measures under the environmental impact mitigation rules were to involve inner-city brownfield sites (information from the



Urban Redevelopment and Quality of Open Spaces. Figure 6

Temporary use of former brownfield as green spaces in the east of Leipzig (Photo: J. Mathey)

Leipzig Office of the Environment). There is now an intermunicipal compensatory land pool which included most of the land that comes into question. However, as in other cities, it is difficult to find areas for permanent planting in Leipzig.

Permanent Downzoning to Farmland or Forest

Urban and peri-urban agriculture has a long tradition [33]. Regardless of the type and intensity of farming, agricultural areas also provide city dwellers with a wide range of recreational and other amenities [34]. The discussion on the greater use of renewable energies and the growing demand for biomass production can be an economically attractive after-use option for clearance sites in the city. An ongoing Experimental Housing and Urban Development study is examining the use of urban open spaces for renewable energy [35]. Examples of agricultural uses are the short rotation plantations in Halle-Neustadt and the Green Bow in Leipzig-Paunsdorf, the layout of which took account of nature conservation aspects and recreational needs. The types of farming feasible on previously developed urban land will depend on the size of the areas, their location in the urban structure, and on locational

prerequisites (e.g., soil quality), but not least on what actors can be found to manage them.

A frequently envisaged category of after-use in renaturation projects is forest or woodland. Forest does not necessarily require the reallocation of holdings, and ensures secure, long-term management, generally by the forest authorities [36]. In contrast to use as conventional green spaces, managing land under the terms of the Forests Act involves much lower investment and maintenance costs, less stringent traffic safety obligations than for the more usual urban open spaces (problem of liability), much lower land values, but also new use options, since public access is maintained. In keeping with the principle of downsizing from the periphery inward, conversion into forest is envisaged primarily on the urban outskirts, with the aim of establishing interconnectivity with the surrounding countryside [37, 38]. In Schwedt “Am Waldrand,” for example, the afforestation of clearance sites is planned; in Weißwasser-Süd forest-like mixed plantations have been established, which after 4 years of development care are to be transferred to the forestry authorities. The “Waldstadt” in Halle-Silberhöhe is being laid out on the periphery of the large housing estate, bordering on an extensive young-stand, near-natural deciduous



Urban Redevelopment and Quality of Open Spaces. Figure 7

Farmland combined with nature conservation on a former military area in Leipzig-Paunsdorf (Photo: Municipality of Leipzig)

forest [36]. Woodland is used in the inner-city to cover gap sites, maintaining the illusion of density and compactness (e.g., “Dunkler Wald” in Leipzig). Such forested areas attain spatial significance only after a relatively long period, and only then do they correspond to the accustomed image, which can initially lead to problems of public acceptance (Figs. 7 and 8).

Experiment “Wilderness in the City”

In many cases, however, the development of brownfield sites proceeds unsystematically. Depending on how long a vacant plot remains undisturbed or on the way in which it is used, different stages of vegetation development with the associated animal species will predominate. This often results in ecologically valuable habitat mosaics [39–41], which can be precious recreation areas and offer opportunities to experience nature [42]. The question of wilderness in the city is much discussed in the debate on urban redevelopment. This is, however, to be understood not so much as a “free succession” of vegetation with land being completely left to its own devices but as a “regulated succession,” a mix of planned, partly cultivated, and “wild” areas [43]. The overgrowth of vacant sites is



Urban Redevelopment and Quality of Open Spaces. Figure 8

“Dark Forest” on a former brownfield in Leipzig (Photo: I. Hartmann)

accepted only if it remains in the prescribed limits and fits into the pattern of new open spaces [44]. “Orderly succession” constitutes a many-faceted habitat mosaic that also exhibits a certain order. This new aesthetic is more acceptable than impenetrable



Urban Redevelopment and Quality of Open Spaces. Figure 9
 “Orderly Succession” on a former military area in Dresden-Nickern (Photo: R. Bendner)

wilderness. Moreover, in comparison with conventional green spaces, considerable savings can be made on layout and maintenance. In contrast to the substantial attention that this subject attracts among the public and experts, it has so far played hardly any role at all in urban redevelopment, and there are very few examples of projects. On a former military site in Dresden-Nickern, a sort of “ruderal park” has been developed, retaining existing spontaneous vegetation, which includes wilderness and natural succession. This ruderal park is accepted and well used by the residents, however, often the development value and amenity quality of such areas is criticized and public acceptance is low ([26], p. 73f.; [45]) (Fig. 9).

Conclusion: New Open Space Qualities Through Urban Redevelopment?

Urban redevelopment has opened up a new field of experimentation and opportunity in the design and use of inner-city open spaces. The new possibilities and opportunities face a complex of obstacles, primarily financial but also legal in nature. They cannot be overcome by traditional strategies: new forms of cooperation are needed. This is also true of coping with the changing demands made on the use of open spaces.

Overall, a change of awareness is required, a shift from growth toward a novel combination of growth and shrinkage. Approaches directed toward a subdivided or fragmented urban structure would appear to be more suitable in handling the new open spaces [46].

The advantages offered by “new greenery” such as improved open space amenity, the concomitant ecological upgrading, or the attention paid to the interests of nature conservation are counterbalanced by a number of disadvantages. One is the fragmentation of the physical urban fabric, the perforation of the city, and, not seldom, the limited use capability of the new open spaces. What is more, these new open spaces are often still encumbered by the remains of urban infrastructure. Sometimes they have a derelict aspect, which contributes to stigmatization. Although urban redevelopment produces many new open and green spaces, it leads to a loss of design quality and amenity when measured against existing standards and the associated demands made of urban open space. The importance of open space planning in shrinking cities is therefore increasing; it is becoming one of the “most important tools of urban development” ([19], p. 648).

In the course of urban redevelopment, new open spaces are also expected to perform urban development

tasks and functions. It is questionable whether they are suitable, for the new open space qualities envisioned by integrated urban development concepts do not emerge at once. To begin with, much takes the appearance of isolated pieces of a puzzle or tiles in a mosaic without recognizable outline, or is provisional and transitory in nature. For example, it is not yet clear what is to become of the experiments with vegetation succession. Many experimental projects – including forms of extensive maintenance, wilderness, and succession – are experienced as impoverishment [47]. In the short and medium term, many cities will have to live with open spaces that attain their aesthetic qualities only gradually ([26], p. 74). Since planning practice largely stands by the accustomed types of open space [46, 48], the permanent conversion of redundant building land into urban green spaces of conventional design and function (e.g., parks, playgrounds and recreation facilities, and allotment gardens) is, despite financial restrictions, likely to remain the most frequent after-use for clearance sites in the inner-city and residential areas. However, as budgets shrink, urban parks departments are already scarcely able to maintain their green areas, let alone take on additional cleared sites. In the future, new types of open spaces are therefore likely to be established alongside the accustomed ones, especially areas that allow stronger forms of succession. Thus a growing number of cities undertaking urban redevelopment can be expected to embark on the adventure of succession ([49], p. 12), but they will by no means be overgrown as a result. One outcome of urban redevelopment in the cities concerned will be a more self-stabilizing nature. In view of the far-reaching transformations that redevelopment cause in the cityscape, however, these changes can be regarded as secondary. In sum, urban redevelopment and more greenery will not directly enhance quality in the conventional sense. The population will have to get used to a new quality of the city and its new open spaces. And the decisive innovation will perhaps be a shift in perspective that allows open spaces to be seen with new eyes.

Outlook: Future Directions

As pointed out (in the conclusions) above, the reintegration of brownfields into the urban fabric by developing green and open spaces (renaturation) is

a big challenge, which in future times will become of more and more importance. There is a chance to reconstruct the urban structure (urban fabric) and give cities new faces, but new overall strategies are necessary and a lot of obstacles have to be overcome. In the planning, practice cities have to face a lot of implementation problems such as (1) availability of brownfields (sites not in the ownership of the community), (2) financing of measurements and maintenance, (3) missing cooperation of different city departments, (4) meeting the needs of residents (public relation, participation), (5) considering legal and strategic requirements, and (6) necessitating the profit gain out of the sites. Finding solutions demands for interdisciplinary (research) approaches, where ecological, social, economical, and planning aspects are taken into consideration to develop overall strategies (and concepts) that lead to a “new urban landscape” with valuable, usable, and accepted green and open spaces in a sustainable urban structure.

Bibliography

Primary Literature

1. Lampen A, Owzar A (2008) *Schrumpfende Städte. Ein Phänomen zwischen Antike und Moderne*. Böhlau, Köln
2. Oswalt P, Reiniets T (2006) *Atlas of shrinking cities*. Ostfildern-Ruit, Hatje Cantz
3. Turok I, Mykhnenko V (2007) The trajectories of European cities, 1960–2005. *Cities: Int J Urban Policy Plan* 24:165–182
4. Demuth B, Moorfeld M, Heiland S (2010) Demografischer Wandel und Naturschutz. In: Bundesamt für Naturschutz (BfN) (ed) *Naturschutz und Biologische Vielfalt Heft 88*. Bonn-Bad Godesberg
5. Haase D, Schetke S (2010) Potential of biodiversity and recreation in shrinking cities: contextualization and operationalization. In: Müller N, Werner P, Kelcey JG (eds) *Urban biodiversity and design, Conservation science and practice series*. Wiley-Blackwell, Oxford, pp 518–538
6. BMVBS (Bundesministerium für Verkehr, Bau- und Stadtentwicklung) & Bundestransferstelle Stadtbau Ost (2008) *Evaluierung des Bund-Länder-Programms Stadtbau Ost*. Berlin
7. Rink D (2010) *Schrumpfen als Transformationsproblem. Ursachen und Verlaufsformen von Schrumpfung in Ostdeutschland*. In: Bernt M, Haus M, Robischon T (eds) *Stadtbau und lokale Politik*. Darmstadt, Schader Stiftung pp 58–77
8. Gatzweiler H-P, Meyer K, Milbert A (2003) *Schrumpfende Städte in Deutschland? Fakten und Trends*. In: *Informationen zur Raumentwicklung*, H.10/11, pp 557–574

9. Hannemann C (2003) Schrumpfende Städte in Ostdeutschland – Ursachen und Folgen einer Stadtentwicklung ohne Wirtschaftswachstum. In: Aus Politik und Zeitgeschichte, Beilage zur Zeitschrift DAS PARLAMENT, B 28, S. 3–8
10. Expertenkommission (2001) Wohnungswirtschaftlicher Strukturwandel in den neuen Bundesländern, Bericht, Berlin, p 4
11. Ragnitz J (2006) Demographische Entwicklung in Ostdeutschland. Halle
12. Müller B, Siedentop S (2004) Wachstum und Schrumpfung in Deutschland – Trends, Perspektiven und Herausforderungen für die räumliche Planung und Entwicklung. In: Deutsche Zeitschrift für Kommunalwissenschaften. Jg. 43, H. 1, pp 14–32
13. Siedentop S, Kausch S (2003) Die siedelungsstrukturelle Dimension der Schrumpfung – Räumliche Trends der Bevölkerungsentwicklung in Ostdeutschland nach der Wende. In: Müller B, Siedentop S (eds) Schrumpfung – Neue Herausforderung für die Regionalentwicklung in Sachsen. Sachsen-Anhalt und Thüringen, Hannover, pp 11–24
14. Herfert G (2004) Die ostdeutsche Schrumpfungslandschaft. In: Geographische Rundschau, Issue 2, pp 57–62
15. Häussermann H, Glock B (2004) New trends in urban development and public policy in Eastern Germany: dealing with the vacant housing problem at the local level. Int J Urban Reg Res 28(4):919–929
16. Bernt M (2009) Partnerships for demolition: the governance of urban renewal in East Germany's shrinking cities. Int J Urban Reg Res 33(3):754–769
17. BMVfW (Bundesministerium für Verkehr, Bau- und Wohnungswesen) (2001) Wettbewerb Stadtumbau Ost. Für lebenswerte Städte und attraktives Wohnen, Berlin
18. BMVfS & BBR (Bundesministerium für Verkehr, Bau- und Stadtentwicklung & Bundesamt für Bauwesen und Raumordnung) (2007) 5 Jahre Stadtumbau Ost – Eine Zwischenbilanz. Zweiter Statusbericht der Bundestransferstelle, Bonn
19. Kunz W (2004) Stadträumliche Qualität durch Freiräume. Beispiele für innovative Planungsideen und –methoden aus dem Stadtumbauprozess in Leipzig. In: Informationen zur Raumentwicklung, H. 11–12, pp 645–658
20. Schwartz C (2004) Extensive Freiraumgestaltung als Strategie im Stadtumbau. In: Informationen zur Raumentwicklung, H. 11–12, pp 735–747
21. Roch I, Banse J (2007) Entwicklungsperspektiven für die Lebensqualitäten Wohnen und Freiraumnutzung in den Städten Bonn und Dresden. In: Schrenk M, Popovich VV, Benedikt J (eds) REAL CORP 2007, proceedings, Vienna, 20–23 May 2007, pp 563–578
22. Böhme C (2007) Die "grüne" Stadt – urbane Qualitäten durch Freiraumentwicklung. In: Deutsche Zeitschrift für Kommunalwissenschaften, J. 46, H. 1, pp 5–9
23. Raumordnungsgesetz (ROG) vom 18. August 1997 (BGBl. I S. 2081, 2102), zuletzt geändert durch Artikel 10 des Gesetzes vom 9. Dezember 2006 (BGBl. I S. 2833)
24. Bundesnaturschutzgesetz (BNatSchG) vom 25. März 2002 (BGBl. I S. 1193), zuletzt geändert durch Artikel 3 des Gesetzes vom 10. Mai 2007 (BGBl. I S. 666)
25. BMU (Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (ed.) (2007): Nationale Strategie zur biologischen Vielfalt self-published
26. Rink D (2005) Surrogate nature or wilderness? Social perceptions and notions of nature in an urban context. In: Kowarik I, Körner S (eds) Wild urban woodlands. Berlin/Heidelberg, Springer pp 67–80
27. GreenKeys Team (2008) GreenKeys @ Your City – a guide for urban green quality. Leibniz Institute of Ecological and Regional Development (IOER), Dresden
28. URGE-Team (2004) Making Greener Cities – a practical guide. UFZ-Bericht Nr. 8/2004 (Stadtökologische Forschungen Nr. 37), UFZ Leipzig-Halle
29. Heck A (2005) "STADTHALTEN" – Leipzig. Temporäre Nutzungen in urbanen Räumen. In: Planerin, 01/05, pp 45–47
30. Heck A (2006) Gestattungsvereinbarungen als Handlungsinstrument der Stadtentwicklung. In: DRL (Deutscher Rat für Landespflege) (ed) Freiraumqualitäten in der zukünftigen Stadtentwicklung. Schriftenreihe des DRL Heft 78, pp 111–114
31. Stadt Leipzig (2005) Das Neue Leipzig. Sterneuerung. Neue Freiräume im Leipziger Osten. Leipzig
32. Mathey J, Kochan B, Stutzriemer S (2003) Biodiversität auf städtischen Brachflächen? Planerische Aspekte naturverträglicher Folgenutzungen. In: Bayrisches Landesamt für Umweltschutz (ed) StadtNatur – Bedeutung der Stadt für die Natur und der Natur für die Stadt, pp 47–57
33. Lohrberg F (2000) Urbane Landwirtschaft als Erlebnisraum. In: Garten + Landschaft, Heft 3, pp 9–12
34. Lauinger H (2005) Urbane Landwirtschaft. In: Oswalt P (Hrsg.) Schrumpfende Städte. Band 2: Handlungskonzepte. Ostfildern-Ruit, Hatje Cantz pp 156–164
35. BBR (2007) ExWoSt Programm. Forschungsprojekt "Nutzung städtischer Freiflächen für erneuerbare Energien." Bonn
36. BMVfW & BBR (Bundesministerium für Verkehr, Bau- und Wohnungswesen & Bundesamt für Bauwesen und Raumordnung) (2004) Zwischennutzungen und neue Freiflächen. Städtische Lebensräume der Zukunft, Berlin
37. Kowarik I, Körner S (eds) (2005) Wild urban woodlands. Berlin/Heidelberg, Springer
38. Burckhardt I et al (2008) Urbane Wälder – Ökologische Stadterneuerung durch Anlage urbaner Waldflächen auf innerstädtischen Flächen im Nutzungswandel – ein Beitrag zur Stadtentwicklung. Bonn
39. Wittig R, Zucchi H (eds) (1993) Städtische Brachflächen und ihre Bedeutung aus der Sicht von Ökologie, Umwelterziehung und Planung, Geobotanische Kolloquien, 9. Frankfurt, Verlag Natur & Wissenschaft Harro Hieronimus & Jürgen Schmidt, Solingen
40. Rebele F, Dettmar J (1996) Industriebrachen. Ökologie und Management, Stuttgart
41. Kowarik I (2004) Neue Wildnis, Naturschutz und Gestaltung. In: Garten + Landschaft 2/2004, pp 12–15
42. Rebele F (2003) Was können Brachflächen zur Innenentwicklung beitragen? In: Arlt G, Kowarik I, Mathey J, Rebele F (eds) Urbane Innentwicklung in Ökologie und Planung. IÖR-Schriften, Bd. 39, Dresden, pp 63–74

43. Rink D (2009) Wilderness: the nature of urban shrinkage? The debate on urban restructuring and renaturation in eastern Germany. *Nat Cult* 4(3):275–292
44. Dettmar J (2002) Alternative Wildnis. In: Garten + Landschaft, Heft 5, pp 15–17
45. Heydenreich S (2006) Baulücke? – Zwischennutzungen! Bürgerschaftliche Zwischennutzungen in schrumpfenden Stadtteilen – das Beispiel Solidago e. V., Leipzig Schönefeld. In: Friedrich-Ebert-Stiftung (Hrsg.) *Luxus der Leere – Chancen für den Stadtumbau*, pp 15–21
46. Rößler S (2010) Freiräume in schrumpfenden Städten. Chancen und Grenzen der Freiraumplanung im Stadtumbau, Dresden
47. Mathey J, Rink D (2010) Urban wastelands – a chance for biodiversity in cities? Ecological aspects, social perceptions and acceptance of wilderness by residents. In: Müller N, Werner P, Kelcey JG (eds) *Urban biodiversity and design*, Conservation science and practice series. Wiley-Blackwell, Oxford, pp 406–424
48. Rößler S (2008) Green space development in shrinking cities – opportunities and constraints. In: GreenKeys Team (ed) *Urban green spaces – a key for sustainable cities*. Conference reader for the international conference in Sofia, Bulgaria, 17–18 Apr 2008. Leibniz Institute of Ecological and Regional Development (IOER), Dresden, pp 71–74
49. Kil W (2005) Die neue Leere – eine Chance für's Grün? In: *Stadt + Grün*, H. 9/2005, S. 9–14

Books and Reviews

- BMVBS (Bundesministerium für Verkehr, Bau- und Stadtentwicklung) (2009) *Renaturierung als Strategie nachhaltiger Stadtentwicklung*. Berlin
- BMVBS & BBR (Bundesministerium für Verkehr, Bau- und Stadtentwicklung & Bundesamt für Bauwesen und Raumordnung) (2006) *Stadtumbau Ost – Stand und Perspektiven*. Erster Statusbericht der Bundestransferstelle, Bonn
- Broggi MF (1999) Ist Wildnis schön und “nützlich”? In: Konold W, Böcker R, Hampicke U (Hrsg.) *Handbuch Naturschutz und Landschaftspflege*. Landsberg, Ecomed pp 1–7

UV Effects on Living Organisms

PHILIPP WEIHS¹, ALOIS W. SCHMALWIESER²

GÜNTHER SCHAUBERGER²

¹University of Natural Resources and Applied Life Sciences, Vienna, Austria

²WG Environmental Health, University of Veterinary Medicine, Vienna, Austria

Article Outline

Glossary

Definition of the Subject

Introduction

Biological Effectiveness of UV Radiation

Transmission to the Effective Skin Dose

Biological Effects on Microorganisms

Effects of UV Radiation on Plants

Animals

Future Directions

Bibliography

Glossary

Abiotic An abiotic factor is a feature or process that is not derived from, or caused by, living organisms but that might affect living organisms in their environment.

Biologically effective UV radiation The biologically effective UV radiation is obtained by integrating the biologically effective spectral irradiance over the whole wavelength range.

Biotic A biotic factor is a feature or process derived from, or caused by, living organisms, which might affect other living organisms in their environment.

Chloroplast Chloroplast is the site of photosynthesis. It is distinguished by its green color, caused by the presence of chlorophyll.

Chromophore A chemical group capable of selective light absorption resulting in the coloration of certain organic compounds. In biological molecules that serve to capture or detect light energy, the chromophore is the moiety that causes a conformational change of the molecule when hit by light.

Ecotherm An organism that regulates its body temperature largely by exchanging heat with its surroundings (e.g., fish, reptile, ...).

Eye damage This includes the keratoconjunctivitis (an inflammation of the cornea and conjunctiva, snow blindness), an acute effect, and chronic damage of the eye lens, caused by absorbing UVA radiation.

Flavonoids An umbrella term that designates a wide spectrum of pigments (including anthocyanins, among others).

Genetic damage DNA absorbs UVB radiation and the absorbed energy can break bonds in the DNA. Most of the DNA breakages are repaired by proteins present in the cell nucleus, but unrepaired genetic damage of the DNA can lead to skin cancers.

Health effects In humans and animals, prolonged exposure to solar UV radiation may result in acute and chronic health effects on the skin, eye, and immune system. UVB exposure induces the production of vitamin D in the skin. The majority of positive health effects are related to this vitamin.

Minimal erythema dose (MED) The amount of UV radiation exposure required to cause a just perceptible reddening of the skin.

Mitochondria “Powerhouse” of the cell, which acts as the site for the production of high-energy compounds. These high-energy compounds are the vital energy source for several cellular processes.

Monomer A molecule that can combine with others to form a polymer.

Morphogenesis Biological process of growth, structure, form, or shape.

Phototaxis A response to a directional light stimulus that involves movement of a whole organism. This is advantageous for phototrophic organisms as they can orient themselves most efficiently to receive light for photosynthesis. Phototaxis is called positive if the movement is in the direction of increasing light intensity and negative if the direction is opposite.

Plant growth The rate of increase in weight and size of plant organs, such as leaf, stem, or root.

Skin damage Beside the sunburn (erythema), a well-known acute effect, chronic skin damage is also observed. This includes elastoses (damage of collagen fibers) and, thereby, aging of the skin and skin cancer. (The three most common skin cancers are basal cell cancer, squamous cell cancer, and melanoma, each of which is named after the type of skin cell from which it arises. Melanoma is one of the less common types of skin cancer but causes the majority (75%) of skin cancer-related deaths.)

Spectral weighting function Wavelength dependency of biologic efficiency of UV radiation. The multiplication of spectral irradiance and a spectral weighting function delivers the biologically effective spectral irradiance. The spectral weighting function is unique for each photobiological process.

UV (ultraviolet) radiation Electromagnetic energy in the wavelength range between 200 and 400 nm. In the solar spectrum on Earth’s surface the wavelength range is limited to 280 and 400 nm.

The spectral range is divided in UVC (200–280 nm), UVB (280–315 nm), and UVA (315–400 nm).

Yield Economic product harvested from plants, for example, seed from pods of soybean, roots from carrots (*Daucus carota*), grain from wheat (*Triticum aestivum*), or seed and lint from cotton (*Gossypium hirsutum*).

Definition of the Subject

Since the energy of an electromagnetic photon increases with decreasing wavelength, the biological effectiveness – defined as the capacity of radiation to produce a specific biological effect – also increases with decreasing wavelength. The largest biological effects are therefore generated in the shortest wavelength range (280–400 nm) of the incident solar radiation spectrum, which is referred to as ultraviolet (UV) radiation. Living organisms (i.e., living systems such as animals, plants, and microorganisms) are influenced by solar radiation and especially also by UV radiation. Three important aspects have to be taken into account in an investigation regarding the UV impact on living organisms: First, the damage potential of UV has to be studied; second, the faculty of living organisms to protect themselves and to adapt has to be examined; and third, the environment and its impact on the UV radiation field has to be taken into account.

Introduction

The decrease in the ozone layer, which was discovered in the early 1980s, has raised big concerns among the scientific community and among the decision makers. The ozone layer shields the Earth from the very harmful UV radiation that belongs to the part of the solar spectrum with the largest photon energy. Before the ozone hole discovery, it was well known that UV radiation may lead to sunburn on humans and animals, and it was also known that UV may have a negative impact on the eyes. It was, however, not clear what impact UV would have on the biosphere if the ozone layer was to diminish more. Would the biosphere be able to adapt? Or, conversely, would the expected increase in UV eradicate all life on Earth? A lot of money was therefore invested in research on the impacts of UV on the biosphere: the UV effects on humans, as well as on microorganisms, plants and

animals, were investigated. In this article, the focus is on microorganisms, plants and animals. An overview of the research results obtained in these topics during the last 30 years is given, in the form of a detailed review of the findings on the damage potential of UV, and on the adaptation potential and the UV defense mechanisms of microorganisms, plants and animals.

Biological Effectiveness of UV Radiation

The main concept that is used throughout this article is that of the “dose,” which describes the biologically effective UV radiation reaching the skin, considered as a relevant receiving surface, and taking into account the Grotthus–Draper law (named after the chemists Christian J.D.T. von Grotthus and John W. Draper). This law states that a tissue effect can only occur if radiation is absorbed. However, absorbed radiation does not necessarily cause a biochemical reaction. Absorbed radiation may simply be converted into heat, or it may be re-emitted as light at a different wavelength. The latter phenomenon is called fluorescence.

The spectral sensitivity of these biochemical effects is taken into account by a spectral weighting function,

whose convolution with the incident solar spectrum results in biologically effective radiation (it would be good to introduce the MED here or in next section).

Biologically Effective Irradiance and Dose

The biologically effective irradiance E_{biol} is calculated by a dimensionless biological weighting function σ_λ and the incident spectral irradiance E_λ from:

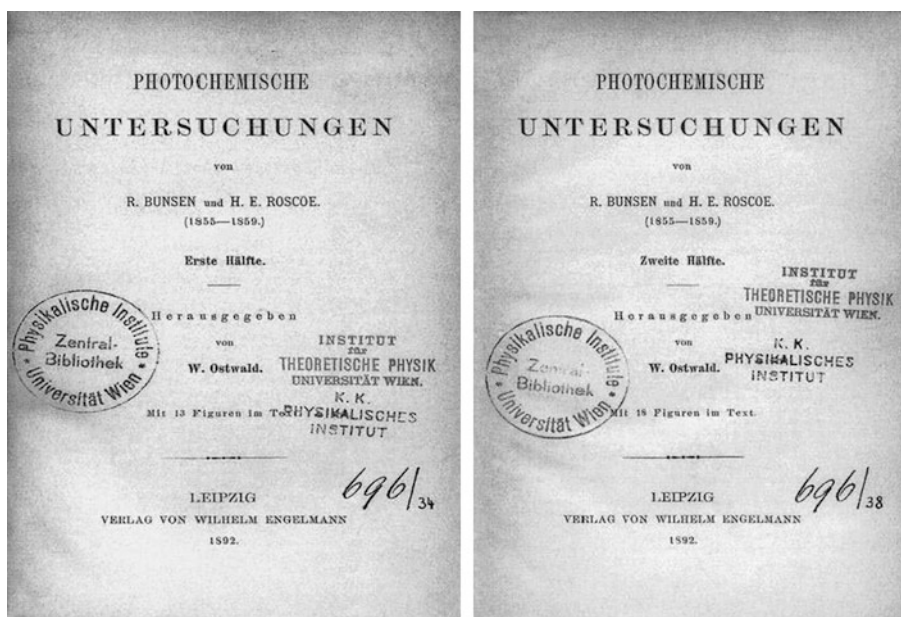
$$E_{\text{biol}} = \int_{280 \text{ nm}}^{400 \text{ nm}} E_\lambda \sigma_\lambda d\lambda. \quad (1)$$

The corresponding biologically effective dose (irradiation) is obtained as the product of the biologically effective irradiance E_{biol} and the time of exposure t :

$$D_{\text{biol}} = E_{\text{biol}} t \quad (2)$$

if it can be assumed that E_{biol} is constant over the period. Otherwise, a summation over time would be necessary.

According to the rule of Bunsen and Roscoe [1] see (Fig. 1), also called reciprocity law, a photochemical reaction is directly proportional to the dose D_{biol} , irrespective of the time t over which this dose is



UV Effects on Living Organisms. Figure 1

Cover page of a re-print of the paper by Bunsen and Roscoe [1], published by Wilhelm Engelmann in the series “Ostwalds Klassiker der modernen Wissenschaft,” Leipzig 1892 [2]

delivered. This often implicit assumption is a simple working hypothesis even though it has been shown that at a constant total dose, the irradiance (intensity) is a major factor that determines both the quality and quantity of the response [2, 3], in the case of UV-induced cell death, photocarcinogenesis, psoralen photochemistry, and effects of low-level laser radiation.

Reciprocity law failures were commonly observed in experiments conducted at either very low or very high radiant fluxes. To account for these failures, Schwarzschild, an astronomer, proposed a modification of the reciprocity law that resulted in a better fit to his low intensity, stellar data. This empirical model later became known as Schwarzschild's law [4] and is given by either [5]

$$D_{\text{biol}} = E_{\text{biol}}^p t \quad (3)$$

or

$$D_{\text{biol}} = E_{\text{biol}} t^p \quad (4)$$

where p is the Schwarzschild coefficient. When $p = 1$, then the Schwarzschild and the reciprocity law are identical. Martin et al. [5] found in a review that, in 11 experiments with erythema, only one deviated from the reciprocity law by $p = 1.2$. For many other biological effects the Schwarzschild coefficient was found to be $p = 1$ for the majority of experiments.

Dose-Response Relationship

A dose-response function (DRF) describes the influence of a biologically effective UV dose, D , on an observed biological effect.

Erythema The erythema is caused by ultraviolet radiation. It can be observed on animals in the same way as on humans. In particular, the dose-response functions and the time course of the reddening of the skin of pigs were investigated in detail.

The dose-response function of skin reddening can be described by a logit function,

$$\Delta EI = \frac{\Delta EI_{\text{max}}}{1 + e^{\frac{\log D - \log D_{0.5}}{-k}}} \quad (5)$$

The erythema index ΔEI is defined via the international color system, $L^*a^*b^*$. The ΔEI value is the difference of the red color value a^* before and after UV

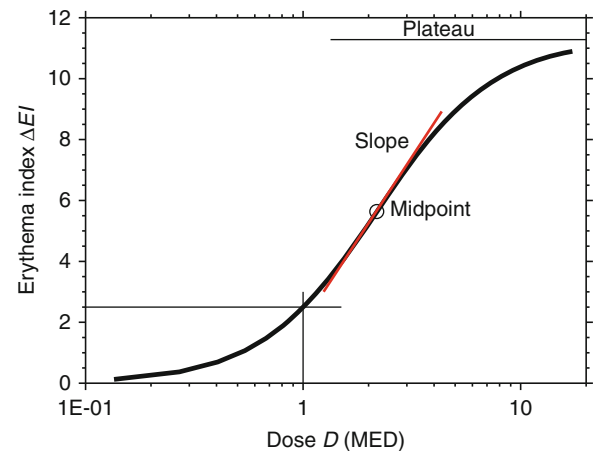
exposure, the UV Dose D (J/m^2), the plateau of the dose-response function (maximum value of the erythema index) ΔEI_{max} , the dose of the midpoint of the function $D_{0.5}$ and the slope of the logit function at the midpoint k (Fig. 2). The measurements of the redness of the skin can be done by a color meter (e.g., Minolta Chroma Meter CR 300) that uses the international color system $L^*a^*b^*$.

Most photobiological effects do not appear immediately after or during exposure but after a certain latent period. For certain effects, the latent period may be in the order of decades (e.g., skin cancer in human), while for other effects, the period is rather short (e.g., erythema).

The visibility of an effect may depend also on time. In a first phase the visibility of an effect increases until reaching a maximum. Afterward the visibility decreases. If an effect is reversible, the visibility of the effect goes back to zero; if the effect is irreversible, it stays visible. Repair mechanisms (photorepair) of the organisms are crucial to make the effect disappear. Sustained irradiation can be additive and enhances the remaining effect (cumulative effect).

The temporal course of an effect may depend on dose and wavelength.

The dose-response function, as well as the time course of the erythema in pigs, were investigated by

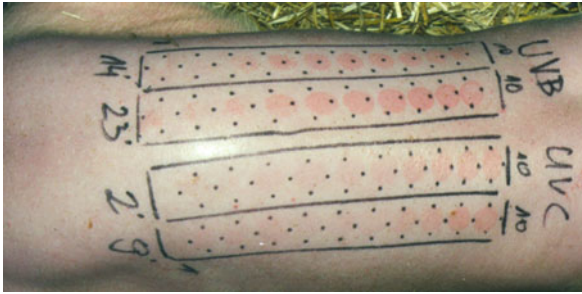


UV Effects on Living Organisms. Figure 2

UVB-caused dose-response curve, 24 h after exposure, for a pig with a minimal erythema dose of $\text{MED} = 148 \text{ J}/\text{m}^2$. In this specific case, the plateau ΔEI_{max} equals 11.28, the midpoint $D_{0.5}$ is 2.75 MED, and the slope k is 0.269

Reischl et al. [6] (Fig. 3). In Fig. 4, the normalized dose-response function of ten pigs (crossbreeding of “Edelschwein” and “Landrasse”) show similar behavior. The differences of the plateau and the slope of the function can be explained by individual sensitivity, similar to the four skin types of humans. The median of the MED was calculated as 165 J/m^2 , the lower quartile 143 J/m^2 , and the upper quartile 184 J/m^2 . In Table 1 the minimal erythema dose (MED) is summarized for cattle, horse and pig.

To describe the time course of the erythema index ΔEI , the doses were divided into three groups: low doses if below 1 MED (sub-erythematous doses), medium doses between 1 and 7 MED, and high doses above 7 MED. The time course of the artificial UVB and UVC exposure is depicted in Fig. 5.



UV Effects on Living Organisms. Figure 3
Erythema (on pig skin?) caused by UVB and UVC radiation [6]

The time course of erythema was bimodal after UVB and UVC exposure. Both spectral ranges show a peak for low doses within about 12 h. Higher doses showed a greater intensity of redness and later fading by UVB than by UVC. Two effects could be distinguished. The first effect caused an early maximum of erythema by low doses, whereas the second effect caused a later, higher maximum by higher doses. These two effects can also be observed in studies on humans [7]. For both spectral ranges, the time course of erythema was primarily related to the UV dose and not to the spectral range.

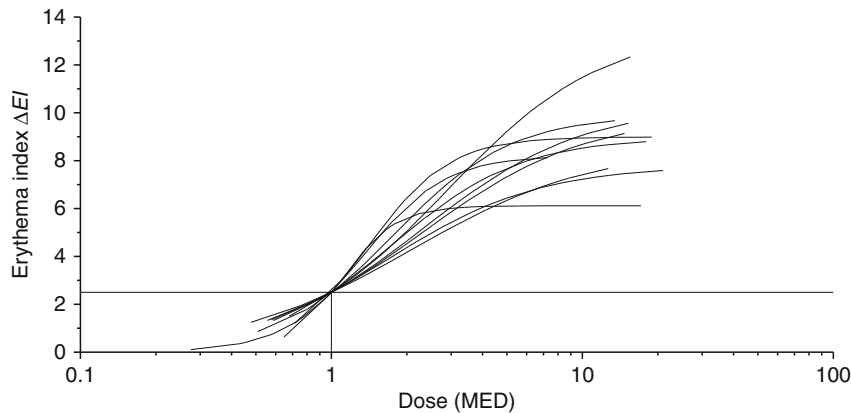
Action Spectra

If one would like to estimate the effectiveness of UV radiation in causing a certain effect, the action spectrum (or biological weighting function) has to be known to weigh the incident spectral irradiance. An action spectrum (AS) describes the spectral efficiency of light in initiating a photochemical or photobiological process.

Conventionally, an AS is determined by exposing a sample to $n(\lambda)$ photons at wavelength λ to produce the effect. Then the wavelength and photon number are varied in such a way that the same effect is produced at all wavelengths. The AS s is then

$$s(\lambda) = C/n(\lambda) \quad (6)$$

where C is a constant.



UV Effects on Living Organisms. Figure 4
Dose-response function for pigs, normalized by the individual minimal erythema dose (MED). The MED was determined by an erythema index $\Delta EI = 2.5$, which is the limit of the visual detection 24 h after a UVB exposure [6]

UV Effects on Living Organisms. Table 1 Minimal erythema dose (MED) for several species

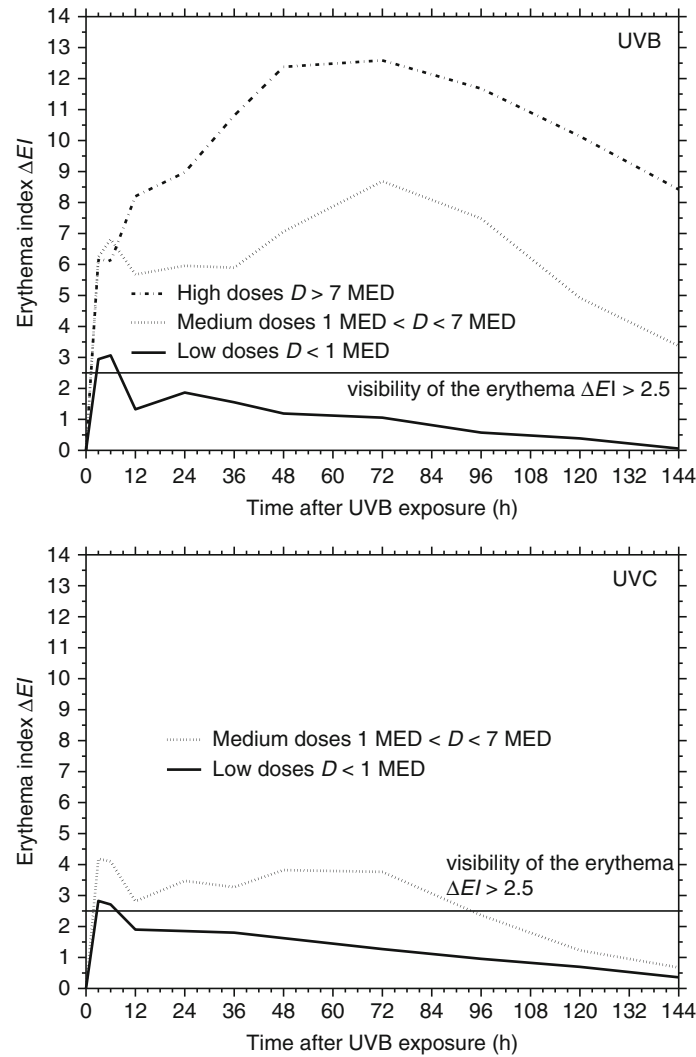
Specie	Minimal erythema dose (MED) (J/m ²)	Author
Cattle	100	Mehlhorn and Steiger [8] ^a
Horse	450	Kasper [9] ^b
Pig	165	Reischl et al. [6] ^b

^aSpectral weighting function not clearly defined
^bReference erythema weighting function by MacKinley and Diffey [10]

The number of photons is called radiant exposure or dose. It is the product of irradiance and the duration of irradiation.

If only one chromophore is present and if this chromophore is in an unbound and monomeric state, the action spectrum will have a shape identical to that of the absorption spectrum of the chromophore. In samples with many absorbing molecules, action spectroscopy is a powerful tool used to identify the specific chromophore for the process under scrutiny.

If free monomeric chromophore molecules are present together with bound or aggregated molecules, the



UV Effects on Living Organisms. Figure 5

Time course of the erythema intensity caused by UVB and UVC radiation, measured by the erythema index ΔEI for three dose levels [6]

situation is more complicated. Aggregated or bound molecules usually have distorted or shifted spectra as compared with monomeric molecules, but are usually less efficient in initiating photochemical processes.

If different chromophores are involved, the AS is a superposition of several absorption spectra.

Action spectra are specific to each process. They may obviously differ, as in the cases of DNA damage, Erythema, and Pigmentation see (Figs. 6–8), depending on the absorbing molecules.

Comparisons between experimentally determined action spectra and absorption spectra of appropriate molecules can sometimes give insight into which molecule is primarily responsible for the effect. Using the bactericidal AS derived by Gates [11] see (Fig. 6), it becomes possible to identify DNA as the genetic material [11, 12]. This observation was confirmed with other unicellular organisms and led to the realization that nucleic acids have a fundamental role in ultraviolet photobiology.

Although it is still possible to glean more information about the photoresponses of organisms that do not meet the criteria above, it may be impossible to identify either the molecule(s) or the mechanism(s) involved in the process being studied. It may not be possible to circumvent these limitations if a complex process, such as plant “growth,” is to be analyzed. An action spectrum for this latter process may be nothing more than a determination of effects as a function of

wavelength, but the exact mechanism causing the effect might remain unknown. Even in this case, such action spectra can be useful in estimating the response to ambient of changing light exposures.

After the discovery of infrared [13, 14] and ultraviolet radiation [15], the crude action spectroscopy was done by dividing whether a certain photobiological effect is caused by UV, or rather visible or infrared radiation. During the nineteenth century action spectroscopy was refined by using colored glass filters to select wavelength ranges. This technique has already allowed identifying chlorophyll as the chromophore most responsible for the growth of plants [16–18].

The first spectral resolved AS (Fig. 6) was reported by Hausser and Vahle [19] after constructing a high-quality quartz-prism monochromator. They first reported the erythema action spectrum, which showed a major peak of activity at 297 nm. The general form of this curve was confirmed by other investigators in the following 10 years [20–22]. Then, in 1934, Coblentz and Stair proposed a “standard erythema curve,” which was subsequently adopted by the Commission Internationale de l’Eclairage [23] in 1935. This was the first standardized AS, which maintained its status until 1987 [24].

In the 1930s, a variety of other important ASs were derived, like for the killing of bacteria [25] and bacteriophage [26], vitamin D synthesis (healing rickets) [27], pigmentation of human skin [28] (Fig. 6), or cell division [29].

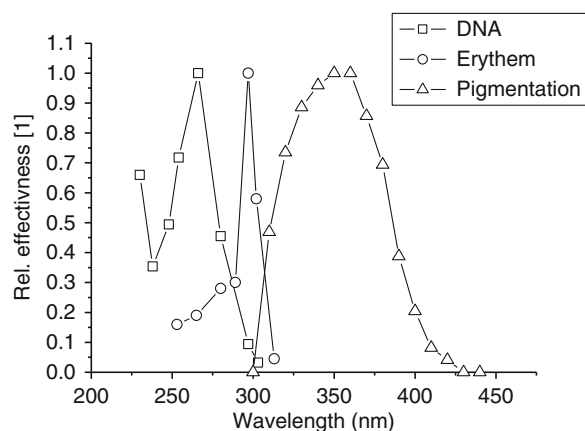
Most of the ASs were derived in the 1980s after the ozone depletion issue became known, to the effect that biologically effective radiation became a topic of interest.

Transmission to the Effective Skin Dose

The estimation of the effective skin or surface dose needs knowledge of the orientation of the receiver in respect to the sun, of the nearby UV environment (reflection, scattering, sky obstruction, ...) and of the protection of the hair coat and behavior (Fig. 7).

Orientation and UV Environment

Measured or calculated solar UV data incident on horizontal surfaces are commonly used to assess



UV Effects on Living Organisms. Figure 6

Early known action spectra such as DNA damage [11], erythema in humans [19], and pigmentation [28]



UV Effects on Living Organisms. Figure 7
3D model of an animal and its environment

the risk to a biological system from the local UV environment. Usually, standard solar radiation measurements are performed for horizontal surfaces and include the incoming irradiance from the whole sky. This incident global irradiance on a horizontal surface is the sum of direct beam and scattered diffuse radiation.

In reality, few biological surfaces are horizontal or flat. However, the surface of an organism can be decomposed into many small inclined planes or “facets.”

Figure 8 depicts such a surface mesh for a cow, along with the corresponding frequency distribution of the tilt of these surface patches. This is convenient for an immobile subject. Obviously, the task of assessing the radiant exposure becomes much more complex when the subject is moving.

The incident irradiance on inclined planes (e.g., body parts of an animal) includes the sum of three radiation components see (Fig. 9):

- Direct beam irradiance: The intensity of the incident direct beam on an inclined plane depends not only on the intensity of the direct beam irradiance but also on the incidence angle between the sun and the normal to the plane.

- Diffuse sky irradiance: The inclined plane does not “see” the whole hemisphere, and the diffuse irradiance is therefore reduced compared to a horizontal receiver.
- Ground-reflected radiation toward the receiver: This reflected radiation depends on the illumination of the reflecting facet (incident irradiance on ground), on the solid angle of the facet, on the reflectance (albedo) (Table 2) of the facet, and of course on the inclination of the receiving plane.

Direct Beam Radiation The direct beam, I , on a given plane depends on the initial intensity of the direct solar beam irradiance incident on a normal plane (I_0), which can be measured or calculated, and on the angle of incidence, β , which is the angle between the solar beam and the surface normal (also called illumination angle)

$$I = I_0 \cos \beta \quad (7)$$

β is calculated using the following equation:

$$\cos \beta_i = \cos \theta_s \cos \theta_{n,i} + \sin \theta_s \sin \theta_{n,i} \cos(\varphi_s - \varphi_{n,i}) \quad (8)$$

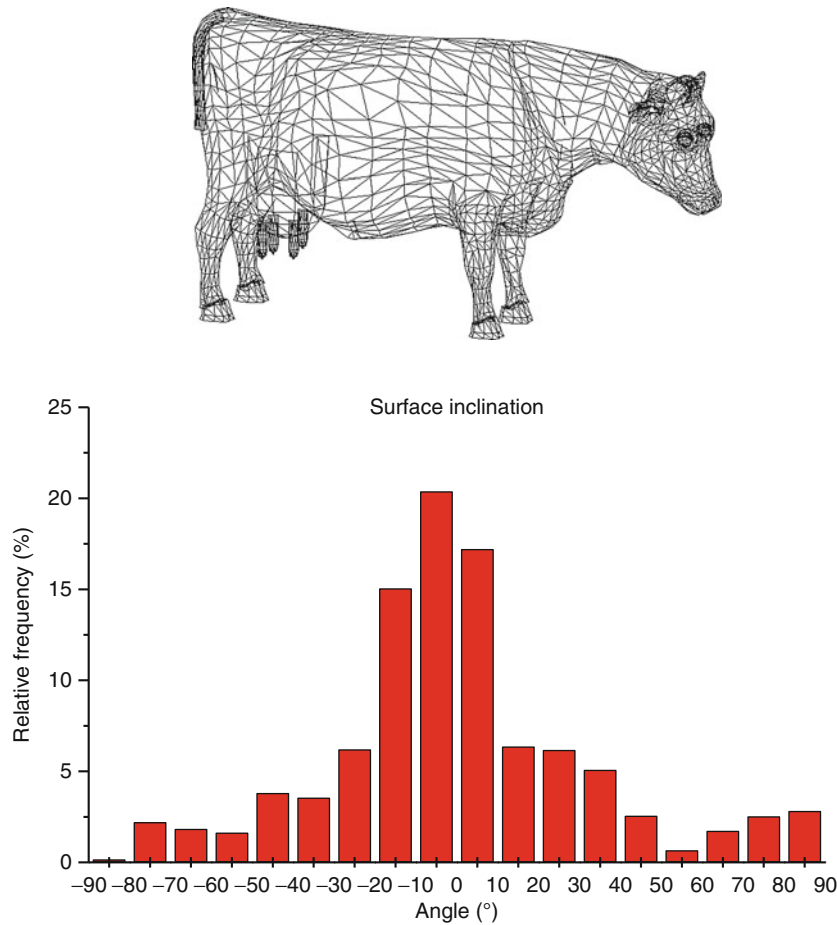
where θ_s is the solar zenith angle, $\theta_{n,i}$ the slope inclination (or tilt angle), $\varphi_{n,i}$ the azimuth (orientation) of the slope, and φ_s the solar azimuth.

Diffuse Irradiance The diffuse irradiance D incident on a horizontal plane is the integral of the sky radiance I over the whole sky hemisphere (please replace I here by something else to avoid confusion since it is already used in Eq. 7).

$$D = \int_0^{\pi/2} \int_0^{2\pi} I \sin \theta \cos \theta \, d\theta d\gamma \quad (9)$$

where θ is the zenith angle and γ the azimuth angle of a sky element.

For the calculation of the incident radiation on an inclined plane, the sky radiance has to be integrated only over the portion of sky seen from the plane. The diffuse irradiance on a given plane may be approximated, assuming an isotropic hemispherical sky radiance distribution, using the following equation [31, 32]:



UV Effects on Living Organisms. Figure 8

3D model of a cow and frequency distribution of surface tilt (0° denoted vertically, +90° horizontally face up and −90° face down)

$$Dt = (1 + \cos(\theta_{n,i}))D/2 \quad (10)$$

where $\theta_{n,i}$ is the tilt angle of the plane, and D is the diffuse horizontal irradiance (Eq. 9). Equation 10 does, however, not take into account the reflected irradiance emanating from the ground, which should be added to the quantity Dt .

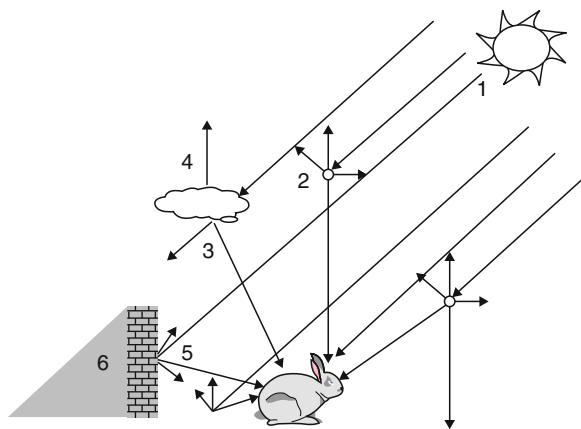
Measurements showed that the ratio of the irradiance on a horizontal plane to the irradiance on the vertical plane has a strong seasonal course, resulting in higher irradiance on the horizontal plane in summer and higher irradiance on the vertical plane in winter at mid- to high-latitudes. The irradiance on vertical planes in winter may be up to 2.5 times higher than in summer [33]. In summer, the ratio of irradiance incident on vertical to irradiance incident on

horizontal planes may be as low as 0.2 for solar zenith angles around 45°. Over snow surfaces the incident irradiance can be much higher [34–36]. Figure 10 shows the influence of the UV albedo on the UV irradiance incident on an inclined plane.

Obstruction of the Sky The obstruction of the horizon shields a sensor/subject from a part of the diffuse irradiance. The surroundings, however, also reflect some radiation toward the sensor or subject.

The obstruction of the horizon may be calculated by integrating the shadow line over all azimuth angles [32]:

$$\text{Obstr} = \int d\alpha d\varphi \quad (11)$$



UV Effects on Living Organisms. Figure 9

Overview of the various radiation components and phenomena that have an impact on the radiation exposure of an animal. 1 direct beam irradiance, 2 radiation scattered by molecules and aerosols, 3 radiation scattered by clouds. (2 and 3 make up the diffuse radiation component), 4 radiation reflected by clouds, 5 reflection by the surroundings, 6 shading (here, by a wall)

UV Effects on Living Organisms. Table 2 Albedo of various surfaces in the UV and visible range

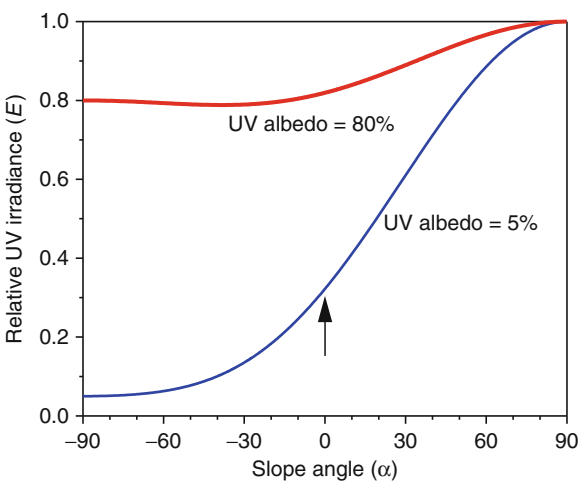
Surface	UV	Visible
Woodland	0.07	0.2
Grass	0.05	0.18
Bare soil	0.1	0.3
New, deep snow	0.9	0.8

Source: Blumthaler and Ambach [30]

where α is the elevation of the object over the horizon and φ its azimuth.

Obstr is the angular obstruction of the horizon (due to the surrounding obstacles, such as mountains) in addition to the obstruction already occurring due to the inclination of the sensor.

By taking this obstruction of the horizon into account, the effective diffuse irradiance Dt on a given plane may be calculated under the assumption of isotropic sky radiance distribution and no reflected irradiance, by using the following equation [31, 32], which is a modification of Eq. 10:



UV Effects on Living Organisms. Figure 10

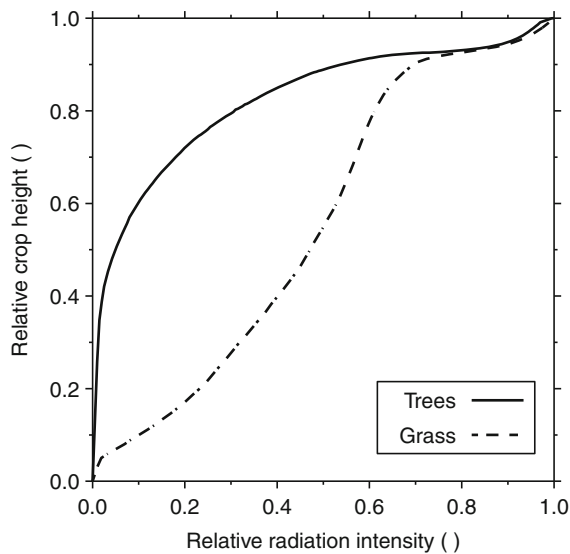
Relative irradiance on a sloped receiving surface in dependence of slope angle and UV albedo

$$Dt = (1 + \cos(\theta_{n,i}))D(1 - \text{Obstr})/2 \tag{12}$$

where Obstr is obtained from Eq. 11.

Models Using Eqs. 7 and 8, the calculation of the incident beam UV radiation on inclined planes is straightforward. Equations 10 and 12, for the determination of the diffuse irradiance on incident planes, bear some uncertainty, because the sky radiance is not perfectly isotropic. Several geometrical models [37–41] for the calculation of the broadband (rather than UV) diffuse irradiance which account also for the non-isotropic component of the sky radiance are available [42]. (Note that [37, 39–41] deal with broadband irradiance, not UV; in the UV, the sky radiance is more isotropic than at longer wavelengths, so that these models should not be used in the present situation . . .) Models reproduce adequately the incident UV radiation on planes facing the equator with a mean bias deviation lower than 4.5%. For planes with other orientations, a larger bias may be found, for example, 10% or more. In recent years, radiative transfer models were adapted to calculate the incident UV on inclined planes [43], taking also into account the obstruction of the horizon [44].

Vegetation In addition to direct shading, vegetation can also be used by animals as protection against UV



UV Effects on Living Organisms. Figure 11

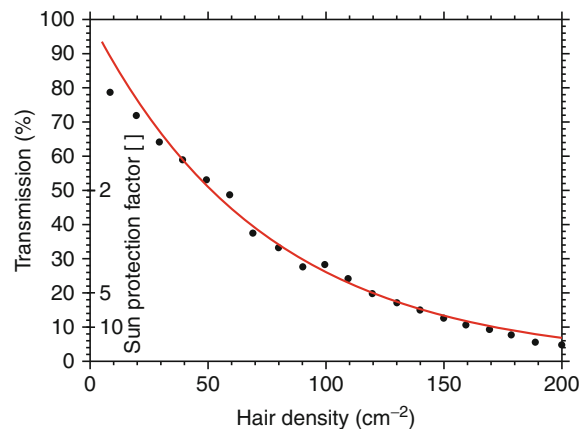
Effect of the relative crop height of trees and grass on the relative radiation intensity normalized to unity by the intensity above the crop, according to Dirmhirn [45]

radiation. The attenuation of radiation results from shadowing by leaves, branches, or stems. Therefore, the geometrical structure of vegetation affects its UV attenuation and protection capacity (Fig. 11). Moreover, the protection is dependent on the seasonal changes like height of vegetation, density, and leaf area [45, 46]. Figure 11 depicts the influence of the relative crop height of trees and grass on the relative radiation intensity normalized to unity by the intensity above the crop, as modeled by Dirmhirn [45].

Behavior

The biggest problem for estimating the effective skin dose is the consideration of the behavior, because it affects the duration of stay within the vegetation. Many herbivores forage in the open field and afterward go back into higher vegetation for cooling, resting, and cowering. For this no investigations are available. The biggest problem for estimating the effective skin dose is the consideration of the behavior, because it affects the duration of exposition. For this only a few investigations are available especially for ectotherms. [please condense these 2 paragraphs]

In 1997, Manning and Grigg [47] found that, for the freshwater turtle, basking (moving and positioning



UV Effects on Living Organisms. Figure 12

UV Transmission and the resulting UV sun protection factor SPF due to the hair coat of cattle, according to Bianca and Wegmann [54]

the body to maximize sun exposure) lasts longer than its thermoregulation would require. Ferguson et al. [48, 49] have shown that Panther chameleons can behaviorally regulate UV exposure or regulate absorption by changing skin color, to an absorption level at which the body already responds with vitamin D production at low UV levels. Karsten et al. [50] have shown that in dependence of their dietary Vitamin D₃ intake, Panther chameleons behaviorally regulate their exposure to solar UV to ensure an optimal Vitamin D₃ status. For some species of lizards, it was previously known that they possess a vitamin D₃ receptor in the brain [51]. Their retina is sensitive to the UV [52, 53] and may therefore be helpful for sun seeking and predator avoidance.

Hair Coat

One important barrier that keeps the ambient dose below the effective skin dose is the hair cover of animals. In addition to its important role as the primary thermal protection of the body, the hair cover is an effective UV protection.

Investigations on the hair coat of horses [9, 55, 56] found a biologically effective UV radiation transmission factor to the skin below 2%. This transmission corresponds to a sun protection factor (SPF) of approximately 50. In the case of cattle, similar results were obtained [8, 54]. Figure 12 depicts the relation



UV Effects on Living Organisms. Figure 13
Erythema of a newborn pig due to outdoor UV exposure

between transmission and hair density. This leads to the conclusion that, whenever the biological effects of UV radiation are analyzed, the natural protection of the hair coat has to be considered.

The lack of UV protection was described by Chapman et al. [57] for newborn furless sheep and for pigs [58]. In Fig. 13, a pig is shown with a whole-body erythema due to outdoor UV exposure.

Transmission Through Water, Snow, and Ice

Water For aquatic life, the optical properties of the immediate habitat – both atmosphere and water – are of great relevance. At the boundary, the optical properties of the interface between air and water, as well as the spectral reflection properties of the benthic division, are crucial [59]. At the water–air boundary, the index of refraction changes and must be taken into account. Reflection is in the order of 5–7% in the UVB and 6–9% in the UVA (e.g., [60]), where the higher values are measured at lower solar elevations.

The penetration of UV radiation into water depends on the optical properties of water as well as on wavelength. Transmittance in water is much lower than in the atmosphere.

Solutes and salinity decrease transmittance due to increased absorption, but most important is the amount of absorbing and scattering suspended particles in the water. Particulate matter is either of terrigenous (erosion and pulverization) or biological origin

(dissolved organic carbon, pigments such as chlorophyll or pheophytin, etc.). Especially in eutrophic fresh water systems and coastal regions, the transparency is affected strongly by these.

Dissolved organic carbon (DOC) is organic material from plants and animals broken down into such a small size that it is “dissolved” into water. DOC is an umbrella term for thousands of different dissolved compounds. The humic or tannin types of DOC are yellow to black in color and can have a great influence on the water’s visible color.

DOC attenuates UVB effectively [61, 62]. Moderate levels of DOC may decrease the UVB within 5 cm down to only 2% compared to the level at the surface [63].

DOC is photochemically degraded by UV [64] into dissolved inorganic carbon (CO , CO_2 , H_2CO_3 , ...) [65]. Through this process, UV indirectly affects its own transmission.

Other important absorbers are photosynthetic pigments like chlorophyll, pheophytin, and others.

UV is also absorbed and scattered by organisms living in water, such as plankton, algae, and sea grasses, which may build large canopies.

In general, transparency is highest in the clear ocean water [61] and high-latitude (dry valleys) [66, 67] or high alpine lakes [68], and lowest in brown (humic) waters (e.g., [69, 70]). Altogether, the depth for a 10% UV transmission can vary from about dozens of meters in the clearest waters to a few centimeters in brown humic waters.

A multicomponent model developed by Smith and Baker [71, 72] describes the total attenuation as the sum of numerous partial attenuation coefficients from water, DOC, pigments, particulate matter, and undefined residuals:

$$K_{\text{Total}}(\lambda) = K_{\text{Water}}(\lambda) + K_{\text{DOC}}(\lambda) + K_{\text{Pigm}}(\lambda) + K_{\text{PM}}(\lambda) + K_{\text{Residual}}(\lambda) \quad (13)$$

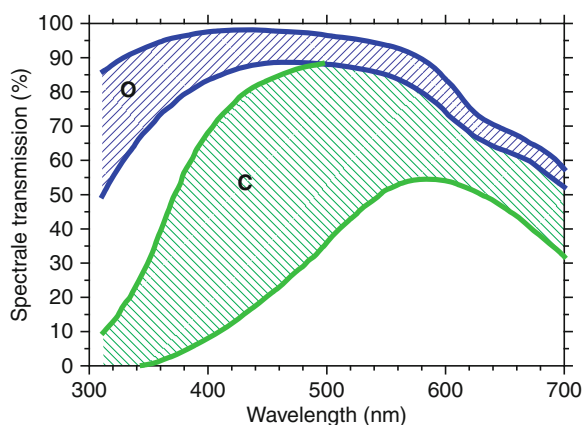
All individual attenuation coefficients on the right-hand side of Eq. 13 depend on wavelength (λ). As a result, UVB is attenuated stronger than UVA and the visible radiation (e.g., [70, 73–75]). In most cases, the main contributor to $K_{\text{Total}}(\lambda)$ is either K_{DOC} (e.g., lowland lakes) or K_{Pigm} (e.g., high alpine lakes) (e.g., [76]).

Once K_{Total} is evaluated, the spectral irradiance E_{λ}^z at a certain depth z can be approximated [77, 78] by:

$$E_{\lambda}^z = -E_{\lambda}^0 \cdot w_{\lambda} \cdot \left(e^{-K_{\text{Total}}(\lambda) \cdot z} - 1 \right) \cdot \frac{1}{K_{\text{Total}}(\lambda)} \quad (14)$$

where w_{λ} is a weighting function for a certain photo-biological effect, and E_{λ}^0 is the spectral irradiance at the surface.

As can be seen from Fig. 14, shorter-wavelength UV radiation is absorbed more effectively than longer-wavelength UV [59, 79].



UV Effects on Living Organisms. Figure 14
Spectral transmission through 1 m of ocean (o) and coastal waters (c)

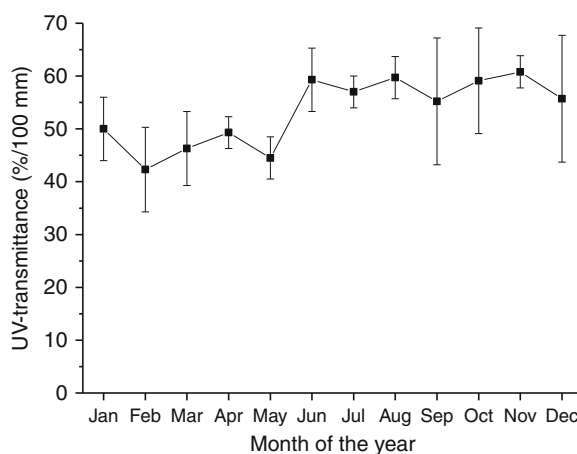
The transparency of water is not constant over time, but shows a temporal (seasonal) variability, which goes hand in hand with the changing amount of solved and dissolved matter in the water (e.g., [64, 69, 80]). As can be seen from Fig. 15, even the transparency of ground-water and well water (drinking water quality) is not constant during the year.

Meteorological factors may also cause variability – snow melting and flood water may transport large amounts of unsolved substances.

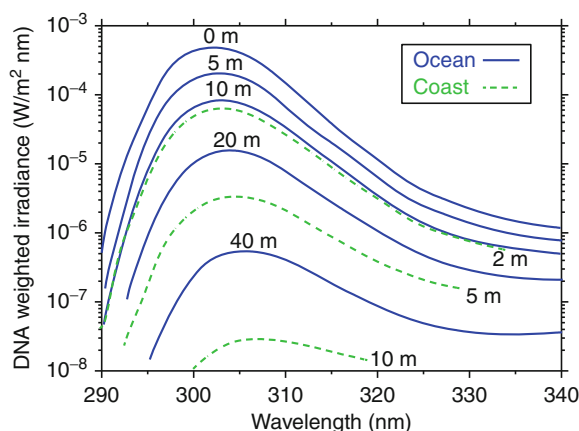
In aquatic regions where water becomes occasionally scarce due, for example, to low water flow or, of course, to the tide cycle, UV radiation on sessile aquatic livings may become extreme, weather permitting.

Smith and Baker [78] have weighted the transmitted UV irradiance with the action spectrum for DNA damage [81]. In Fig. 16, the spectral biological effective radiation for waters with a small amount of soluble organic material and a small content of chlorophyll (clear oceanic water), as well as offshore water with high biological activity, is depicted. The spectral maximum of the 10-m depth line is attenuated by less than an order of magnitude, whereas it is four orders of magnitude for offshore waters.

It can also be seen that the ozone-affected UVB range provides the largest amount of irradiance for DNA damage, although the UVB from the sun, after transmittance through water, has a lower magnitude than UVA.



UV Effects on Living Organisms. Figure 15
Annual course of transmittance (100 mm) of groundwater (drinking water)



UV Effects on Living Organisms. Figure 16

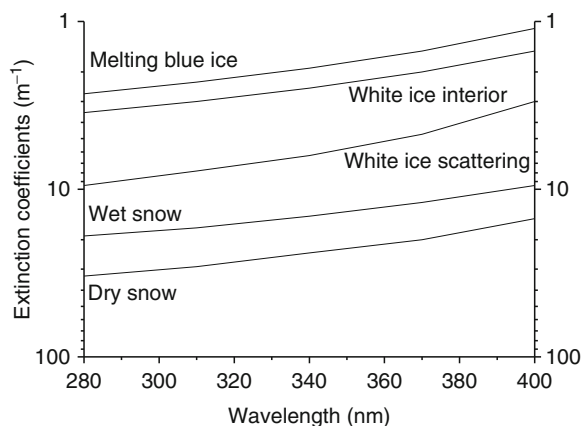
DNA-weighted spectral irradiance in ocean and coastal waters at various depths

Ice and Snow Ice and snow reflect a significant part of the incoming UV radiation. New dry snow may reflect over 90% of the incident UVB. New wet snow reflects over 80%. The amounts reflected by old dry snow and old wet snow are 82% and 74%, respectively [82, 83]. The albedo of snow may change within a day by around ± 0.05 with lower values shortly after noon, which may be caused by a change in the optical properties of snow [84], and by an increase in specularity effects at low sun angles.

Organisms living above the snow cover are therefore at serious risk of UV overexposure because they receive high amount of additional UV. Moreover, the intense reflected UV emanates from below, which represents a completely different direction than usual. This upwelling UV irradiance may hit body parts that are seldom exposed to UV, and are therefore less protected by pigmentation or fur, or less shielded by other body parts. A well-known effect is snow blindness, which humans also experience.

In spite of this high reflectance, a certain amount of UV penetrates into the snow cover (Fig. 17). Nevertheless, snow cover usually provides high UV protection for the organisms living beneath it (see, e.g., [85]). For instance, a snow layer of 5 cm can reduce the UV radiation down to 15–20%, and a layer of 15 cm can reduce it by nearly two orders of magnitudes [85, 86].

The reflectance of ice is somewhat lower than that of snow. It varies from about 30–80%, as a function of



UV Effects on Living Organisms. Figure 17

UV extinction coefficients of ice and snow (After data from [89, 92])

surface structure and temperature [87–89]. The transparency of ice to UV depends on age and clearness of the ice and on wavelength (Fig. 17). UVB is attenuated stronger than UVA or visible radiation. In ice, a layer of 2.5 m can reduce UVA and UVB by up to 2 and 2.5 orders of magnitudes, respectively [90].

One meter of fresh, thin sea ice transmits roughly 20% of the incident UV radiation, whereas cold first-year ice transmits less than 2%. Different layers within the ice, from very white interior ice to almost clear surface ice have different UV extinction coefficients [89, 91, 92].

It was shown by Vincent et al. [67] that the transmitted UV below 3.5 m of very clear ice is still intense enough to inhibit algal growth.

Ice on lakes and rivers is more transparent than that on seawater, because during freeze-up DOC is removed from the ice [93].

Ice is often covered with snow. This combination, of course, affects both albedo and transmittance [87]. A snow cover of ≈ 5 cm (1–9 cm) on top of ≈ 0.9 m (0.5–1.3 m) of ice reduces UVB by 2–13% and UVA by 5–19% [94].

Small changes in snow and ice depth can have an extreme influence on UV radiation. This effect is much stronger than that caused by usual changes in atmospheric parameters like cloudiness or total ozone column [95].

Nowadays, special consideration on the transmittance of ice and snow is needed in the context of

climate change. During shorter periods of snow and ice cover, the “underlying” aquatic or terrestrial organisms can become exposed, even though they are normally shielded from UV radiation. In addition to changing temperatures, a change in water flow can be responsible. An ecologically very important effect is the thinning of ice cover of the Arctic and Antarctic Seas, which leads to an enhanced exposure of a very large aquatic ecosystem.

Ice and snow also provide habitat for microbial assemblages (microalgae, bacteria, protozoa, ...). They can be found in melt ponds on the ice surface, in brine channels within the ice, in slush layers between ice and snow, or between black and white ice, attached to the bottom of the ice or in the sub-ice platelet layer (e.g., [96, 97]). In hypersaline melt ponds above the surface of ice, microbial life can be found at temperatures around -10°C .

On the top of ice and snow, the so-called snow algae (cold-tolerant algae and cyanobacteria) can be found during alpine and polar summers. Algal bloom colors ice and snow with red hues, resulting in what is called red snow or watermelon snow.

Biological Effects on Microorganisms

What are called microorganisms or microbes consist of various types of very small life forms. Microorganisms can be found in the air, in waters, in the ground, on surfaces, and inside other organisms. They can also be found in the harshest environments. For instance, bacteria have been found around hydrothermal vents (hot springs, fumaroles, geysers, or “black smoke” at the sea ground), where water effuses at temperatures up to 350°C . These thermophiles grow best at temperatures above 80°C and accept temperature up to 113°C . At the other extreme, microorganisms can grow in hypersaline lakes at temperatures as low as -10°C . Regarding their tolerance to pressure, both extremes can be accommodated: there are barophilic deep-sea bacteria, and there are also microorganisms in the mesosphere.

Microorganisms (such as viruses, bacteria, fungi, and protozoa) are as manifold and different as one can imagine. Their size may be smaller than the wavelength of light (e.g., 40 nm for viruses). Virus and bacteriophage are the most simple microorganisms.

They consist only of RNA or DNA surrounded by a protein. Both need a host cell for reproduction. Bacteriophages are viruses that infect bacteria. The cells become lysed, which gave bacteriophages their name: “bacteria eaters.”

For humans and animals, a variety of microbes are necessary to stay healthy (e.g., in the gastrointestinal tract, skin, mucous membranes, ...), while other microbes can cause severe diseases (pathogens). The human body consists of approximately 10^{13} cells, but also hosts ten times more microorganisms [98, 99].

Finally, microorganisms play an important role in the biosphere as they build up to 90% of the DNA biomass in marine ecosystems, and play a vital role in the geochemical and global carbon cycle.

UV Environment of Microorganisms

The life span of microorganisms outside a host is defined by temperature, humidity, and solar (UV) radiation. Free moving microorganisms receive UV radiation from all directions. Therefore, the optical properties of the surrounding medium, including its reflectance, are of importance.

On the one hand, solid or organic particulate matter in water or air reduces transmittance of UV, and on the other hand, microorganisms have the ability to adhere on surfaces, so they can make use of particulate matter as radiation protection and vehicles. Clumping of microorganisms is also a strategy of radiation protection in an open environment.

Airborne Microorganisms Microorganisms can be found at all heights in the atmosphere. In the late 1800s and early 1900s, microorganisms were already collected during balloon flights [100, 101]. In 1936, Rogers and Meier [102] collected microorganisms at a height of 20 km. Imshenetsky et al. [103] could even collect reproductive forms of microorganisms (spores of *Cicirnella*, *Penicillium*, *Aspergillus*, and *Papulaspora*, as well as *Micrococcus* and *Mycobacterium*) from the mesosphere between 48 and 77 km (-60°C) using meteorological rockets. Five of the six microorganisms had synthesized pigmentation. At such an altitude, UV radiation is several times higher than on the ground. Moreover, the spectral range is broader, extending to the UVC range, since most of

ozone is below that level, in the stratosphere at about 20–22 km.

Altitude-dependent UV radiation defines the species composition [104]. Bacteria species in the stratosphere are more resistant than in the troposphere or at the ground [105]. Beside pigmentation, clumping and forming aggregates enhances the chances of survival of microorganisms in the stratosphere [105]. In the boundary layer of the atmosphere (i.e., below about 2 km), only around 65% of microorganisms are pigmented [106]. Predominant species are *Staphylococcus*, *Listeria*, and *Micrococcus*.

Small sizes and low density allow microorganisms to remain airborne for long periods before they sediment to the ground. Long-distance transport of fungal spores was observed already in the 1930s [107]. Desert dust enables intercontinental exchanges by offering a UV-shielding vehicle for microorganisms [108–110]. Biomass fires cause strong convection, which is more effective in bringing up microorganisms that are not burned to heights of 3 km than surface stormy winds. Smoke clouds shield them from UV and act as a vehicle over long distances [111]. This explains why, for example, plant fungi can cause allergic reactions or trigger asthma at locations thousands kilometers away from their origin [112].

Airborne microorganisms are influenced strongly by UV. This can be seen by the fact that nonpigmented microorganisms decline strongly with increasing solar UV radiation during the day and during a year. Pigmented microorganisms exhibit a similar pattern but with lower amplitude [113].

Microorganisms released from humans or animals (coughing, sneezing, and talking) into the air have a rather short range (around 12 m) [114] before adhering on surfaces. From these surfaces, they can be re-aerosoled. Survival may last up to 48–72 h. It is assumed that the survival of some microorganisms like influenza virus in air depends less on humidity and temperature than on UV radiation [115, 116].

Waterborne Microorganisms In aquatic environments, bacterioplankton may contribute up to 90% of the total cellular DNA [117–121], and up to 40% of the total planktonic carbon biomass [122]. In addition, bacteria may process up to 50% of algal primary productivity in marine systems [123].

The size of most bacterioplankton cells precludes effective cellular shading or protective pigmentation [124]. Their metabolism is affected by solar UVB down to a depth of 5 m, and by UVA down to 15 m in clear ocean waters [125]. It has been shown that bacterial abundance [126] depends on light cycle as well as on metabolism [127], DNA, protein synthesis, and degradative enzyme activities [128]. Various studies have investigated the effect of UV on different aquatic systems, especially in marine environments [129–132]. Studies on the impact of UV radiation on bacterioplankton have also been carried out in other aquatic systems, such as (alpine) lakes [69, 76, 128, 130, 132–134].

Other authors have done research on biodiversity at extreme UV locations, such as the Himalayas [135–138] and the Andes [139–141]. Moreover, viruses were found in ice-covered Antarctic lakes [142]. The effects of UV in rivers are still poorly investigated.

The most resistant waterborne microorganisms are *Bacillus* sp., *Acinetobacter* sp., *Pseudomonas* sp., *Sphingomonas* sp., *Staphylococcus* sp., and *Stenotrophomonas* sp. These were found in lakes [139, 141] and wetlands [140] up to an altitude of 4,600 m, where UV radiation is approximately 170% that at sea level.

Ecologically and economically important are cyanobacteria (blue-green algae), which are found on wet farmland (rice) all over the world. Cyanobacteria are quite sensitive to UV. Cyanobacteria fix atmospheric nitrogen and are important primary producers in the ocean, but are also found in habitats as diverse as freshwater or hypersaline inland lakes, or arid areas where they are a major component of biological soil crusts.

Effects of UV Radiation on Microorganisms

Damage Microorganisms can be easily damaged by UV radiation. In 1877, Downes and Blunt [143] established that a beneficial effect of sunlight is bactericidal. The experiments that followed revealed that blue/violet is the most effective color to kill bacteria. Hertel [144, 145] could then definitively show that it is the UV part of the spectrum that is responsible. In parallel, it was also found that the bactericidal effect increases with altitude. A detailed review of the early

years of research on the damaging effect by UV is given by Hockberger [146].

Damage can be manifold, and includes growth reduction, decrease of reproduction rate, of metabolism rate, of infectivity, of mobility, or other effects. The main targets for UV damage are RNA, DNA, and proteins.

The wavelength range where UV is most effective in inactivation is quite similar to that where DNA has the highest absorbance (UVC and UVB range) with maximum effectiveness – of course outside the natural UV – i.e., around 265 nm. The range of highest efficiency is also called germicidal range.

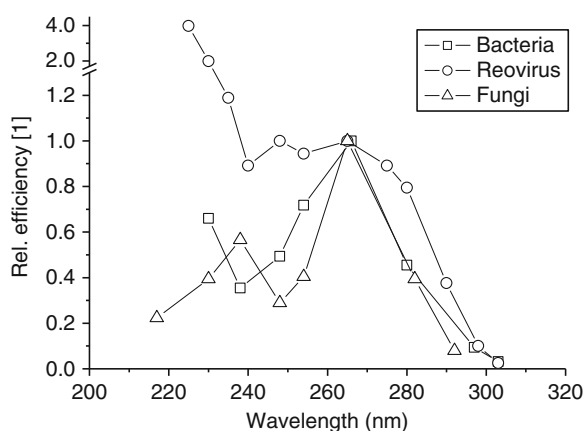
Our knowledge of UV-induced damage results mainly from studies in the UVC and UVB, where efficiency is much lower than in the UVA [147]. However, solar UVA is magnitudes higher than solar UVB. It can thus be assumed that solar UVA contributes also to damage, although only a few action spectra have been proposed for the UVA (Fig. 18).

Disinfection Whereas sterilization results in a total absence of microorganisms, disinfection denotes a substantial reduction of the number of microorganism, by at least four orders of magnitude. Hygiene applications like disinfection make use of UV inactivation [149–152]; inactivation means that the existing microorganisms lose their ability to reproduce. UV

inactivation results from the direct absorption of UV radiation by the microorganism. This absorption triggers an intracellular photochemical reaction that changes the biochemical structure of nucleic acids. These changes may lead to the inhibition of transcription and replication of nucleic acids, thus making the organism sterile and incapable of spreading infection when entering a host.

At the beginning of the twentieth century, Henry et al. [153] found that UVC's disinfection potential was much stronger than UVB or UVA. After developing a mercury vapor quartz lamp and a quartz tube for protecting the lamp against water, the first application of UV irradiation in disinfection of drinking water was realized in Marseilles, France.

Nowadays, disinfection of air, water, and surfaces is generally done by the application of low- or medium-pressure mercury lamps, which emit up to 90% of their light around 254 nm. However, disinfection of drinking water is also possible with solar UV radiation (e.g., Wegelin et al. [154]). The methods can be as simple as using UV-transmitting PET bottles (e.g., McGuigan et al. [155]) or batch reactors (e.g., Navntoft et al. [156]). Serious pathogens like *Salmonella*, *Shigella dysenteriae*, *Escherichia coli*, *Vibrio cholera*, *Pseudomonas aeruginosa*, oocysts of *Cryptosporidium parvum*, cysts of *Giardia muris*, *Candida albicans* (yeast), fungus like *Fusarium solani*, or the Poliovirus can be inactivated effectively by solar radiation. With the current booming interest in solar energy applications, various methods of detoxification and disinfection of water by means of concentrated solar radiation are being successfully experimented, both for small-scale (rural) and industrial-scale operations (e.g., [157–159]).



UV Effects on Living Organisms. Figure 18

Action spectra for inactivation of Bacteria *Escherichia coli* [11], Reovirus-3 [148], and fungi *Trichophyton mentagrophytes* [12]

Protection and Repair The simplest microorganisms have no intrinsic mechanism for UV protection, since their small size does not allow pigmentation or cellular shading. Microorganisms that possess pigments in their outer layer are more resistant. Highly resistant to UV radiation are spores and (oo)cysts. However, UV sensitivity varies a lot, even within a single group, by a factor of up to 10 [160].

Spores and (oo)cysts are reproductive structures that can adapt to dispersal and can survive under unfavorable conditions for extended periods of time.

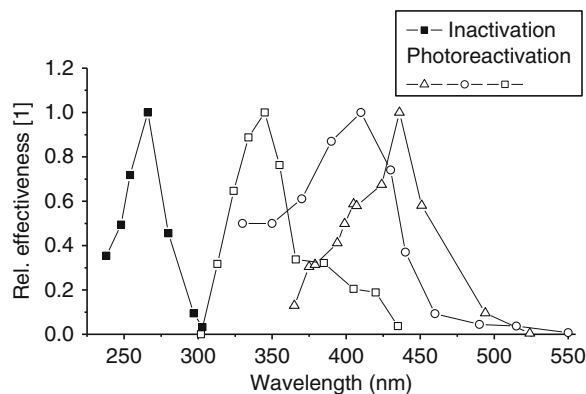
In such a stage, microorganisms have suspended “life,” but are able to start again as soon as the environment becomes more favorable.

Mechanical strategies for radiation protection are clumping of microorganisms, forming aggregates, and building films. Dead microorganisms can act as a shield. Additionally, the biological effectiveness of UV is decreased by orders of magnitude if the targeted (pathogenic/parasitic) microorganism has entered a host.

A strategy for UV protection that microorganisms have elaborated is to postpone UV-sensitive processes until the end of the day or nighttime [161]. Microorganisms exhibit a 24-h cell division cycles, whereas DNA replication and cell division occurs during the night, since all division processes are sensitive to solar UV radiation.

Microorganisms – like all other life forms – know how to defend themselves against damage from UV radiation. There are repair mechanisms that may reduce damage to a certain degree. Nucleic acids, for example, can be repaired in a process termed “photoreactivation” in the presence of light, or “dark repair” in the absence of light. Viruses and phages can take advantage of the reactivity capacity of their host.

The action spectra for photorepair exhibit maxima at three different wavelengths: 345, 400, and 430 nm. Figure 19 depicts the action spectra for damage and photorepair. It can be seen that UVA and blue light are responsible for photorepair.



UV Effects on Living Organisms. Figure 19
Action spectra for DNA damage and photorepair

Since damage is caused by UVB and reactivation by UVA, their ratio is important. This ratio changes during the day, with season, latitude, altitude, total ozone column, and to a certain extent with cloudiness. Low solar radiation, low altitude, high total ozone amount, and high cloudiness tend to favor UVA.

Recent studies have reported an increased UV resistance of environmental bacteria and bacterial spores, compared to lab-grown strains [162]. This means that they could have the ability to adapt to their UV environment.

Vital Effects Damage is not the only effect of UV radiation on microorganisms. In many species of fungi, for instance, UV may also act by igniting new stages in development, like sporulation. Moreover, a variety of morphogenetics depends on UV radiation. Microorganisms can make use of UV for environmental information too. Metabolism, as well as the circadian clock, can also be triggered by UV.

UV modulates activity of bacteria that are responsible for geochemical transformative reactions in the nitrogen and phosphorus cycles [163–165].

Changing UV Environment

As any other organism, microorganisms are well adapted to their environment. However, they have the important advantage to adapt very rapidly to any environmental change, and mutate if necessary.

In general, an increase in UV radiation is detrimental to them, resulting in reduced life span, decreased ability to reproduce, shorter mobility range, reduced habitat, or reduced population.

Still, there are many ways by which microorganisms may profit from environmental or climatic changes. Decreased UV radiation, for example, by enhanced cloudiness, aerosol load or dust, results in decreased damage, which in turn may allow a longer pathway to find a potential host (e.g., [116]). In the case of pathogens, a higher transmission rate and therefore outbreaks of diseases in humans may occur more frequently. For instance, the seasonal outbreak pattern of influenza in many parts of the world is currently believed to result from annual changes in solar UV [166]. Differences in the spring and autumn outbreaks could be explained by different total ozone amounts

and cloudiness, which in turn change the relationship between damage and photoreactivation.

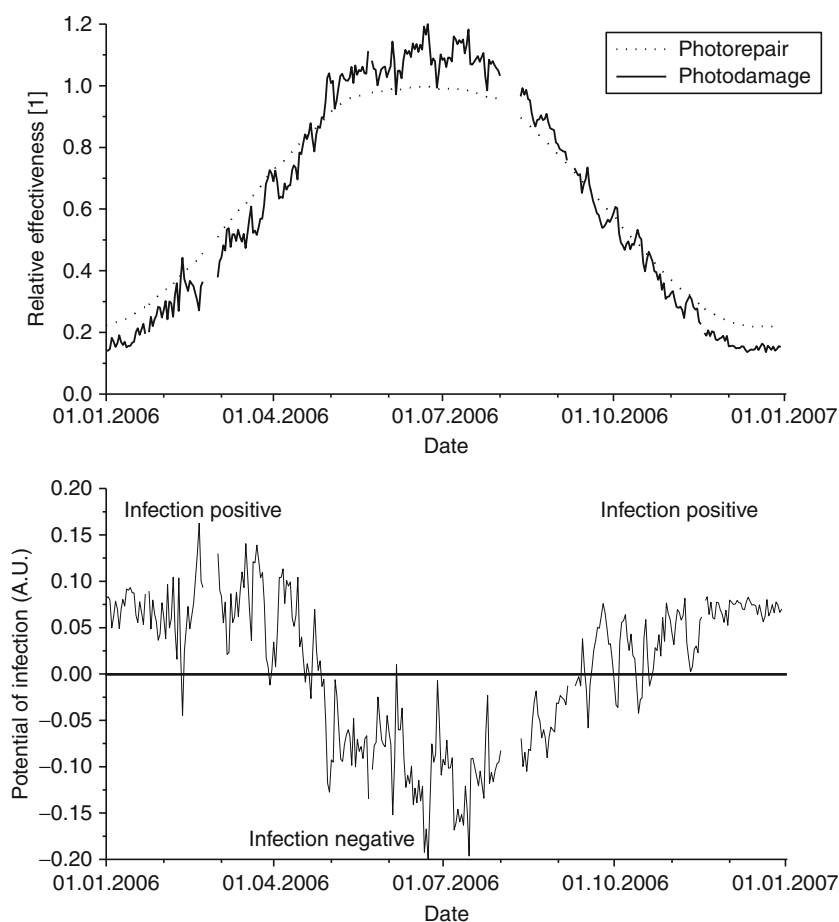
If other environmental parameters change (like temperature or the availability of vectors), microorganisms can settle in a new environment where damage through solar radiation is maybe lower and potential hosts are not prepared for the blight. It has been observed that during the past few years, pathogens have already started to migrate toward higher latitudes.

An important factor controlling infections is the immune response of the host which can be suppressed by solar UV [167]. A lowered immune suppression within a human population enables higher infection rates and more frequent outbreaks. How the

UV-initiated vitamin D level in humans could influence the outbreak of infection diseases is currently being discussed [168] (Fig. 20).

Effects of UV Radiation on Plants

One factor that increases the risk of UVB damage on plants is the fact that they are sessile organisms that are fixed on the ground. UVB damaging effects on plants include the destruction of the cell membranes and all organelles within the cell, including mitochondria, chloroplasts, and deoxyribose nucleic acid (DNA) within the nucleus [169]. These damages to the cell organelles in turn influence the metabolic processes of



UV Effects on Living Organisms. Figure 20

Relative annual course of damaging and reactivating efficiency of solar UV over Vienna, Austria. Calculations are done for clear skies to point out the influence of total ozone

UV Effects on Living Organisms. Table 3 Effects of UVB exposure on various physiological processes in plants

Trait	Decreases	Increases	No effect
DNA damage		X	
Protein destruction		X	
Fatty acid destruction		X	
Photosynthesis	X		
Photosystem I	X		
Photosystem II	X		
Rubisco	X		
Stomata closure		X	
Chlorophylls	X		
Flavonoids		X	
Waxes		X	
Epidermal hairs		X	
Cuticle thickness		X	
Reproduction			
Pollen viability	X		
Pollen tube growth	X		
Fertilization	X		
Cell division	X		
Cell size			X

Source: Prasad et al. [170]

the plant such as respiration, photosynthesis, growth, and reproduction. These UVB damages eventually impact crop yield and quality. The qualitative effects of UVB radiation exposure on the physiological processes, growth, and yield of plants are shown in Tables 3 and 4 [170]. Many factors may influence the extent of the UV impact on plants. The intensity and length of exposure, as well as the type of species and cultivar directly influences the sensitivity of plants to UV.

UVB Radiation Damages

Genetic Damage (DNA Damage) The nucleus of each cell contains deoxyribonucleic acid (DNA), nucleic acid, which contains the genetic instructions used in the development and functioning of plants. DNA is very sensitive to UV radiation [81, 171],

UV Effects on Living Organisms. Table 4 Effects of exposure to UVB radiation on various growth and yield parameters in plants

Trait	Decreases	Increases	No effect
Photosynthesis	Y		
Stomatal conductance	Y		
Phenology			Y
Senescence		Y	
Plant height	Y		
Branching		Y	
Leaf area	Y		
Leaf growth and expansion	Y		
Leaf thickness		Y	
Specific leaf weight		Y	
Dry matter production	Y		
Flowering	Y	Y	
Fruit (grain) number	Y		
Fruit (grain) weight	Y		
Yield	Y		
Quality	Y		
Disease incidence			
Powdery mildew	Y		
Rust		Y	
Insect	Y	Y	

Source: Prasad et al. [170]

which is primarily absorbed in the UVB region by the cell. If DNA is exposed to UVB radiation, effects include (1) breakage of bonds in the DNA and DNA-protein crosslinks, (2) chromosomal aberrations, (3) chromosomal breakage, and (4) exchange and production of toxic and mutagenic photoproducts.

Ultrastructural Cell Damage Ultrastructural changes may occur due to exposure to UV. Among other things, these ultrastructural changes include damage and dilation of the nuclear membrane, rupture of the chloroplast wall, swelling of chloroplasts,

dilation of thylakoids, swollen cisternae in the endoplasmic reticulum, and damage to mitochondria [170].

Inhibition of Photosynthesis Photosynthesis is the process by which plants convert carbon dioxide and water into organic compounds, especially sugars. The photosynthesis apparatus is one important target site of UVB destruction.

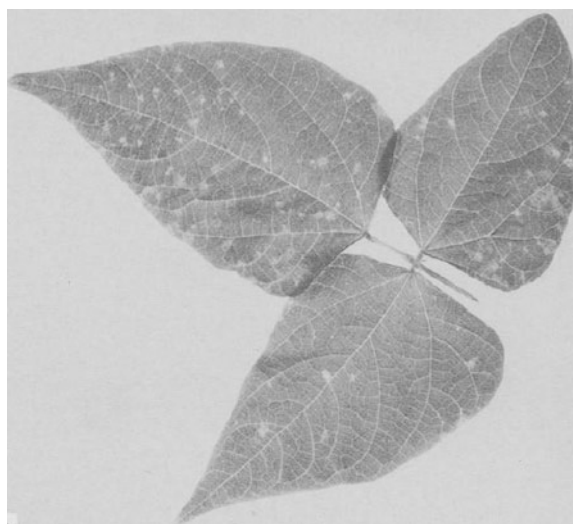
Among the direct effects of UVB that affect photosynthesis are [170]: (1) damage to the ultrastructure of chloroplasts, which are principal sites for photosynthesis; (2) impairment of light energy transfer of photosystem II and to a lower extent photosystem I; (3) reduction of activity of Rubisco (rubisco 1,5-bisphosphate carboxylase-oxygenase); (4) reduction of carbon dioxide fixation and oxygen evolution; and (5) diminution of the starch and chlorophyll content. For field crops [172], a decrease in photosynthesis was observed and was explained both by direct effects (on the photosystem) and indirect effects (decrease in leaf pigments and leaf area).

Plant Morphology and Architecture UVB leads to strong morphological changes in plants [170]. Leaves that are exposed to enhanced UVB radiation first develop irregular chlorotic patches. With a continuation of exposure to UVB, the chlorotic patches (Fig. 21) turn to brown necrotic spots before they later die.

Other effects of UVB are: reduction of the plant height, decrease in individual leaf size (Fig. 22), decrease of tiller number, and reduction of branch length [172]. These morphological changes also lead to a reduction of the density of the canopy, which in turn intercepts less UVB radiation than a plant grown under favorable conditions.

Phenology and Reproductive Processes Neither early bud, flower development, nor the time of first flower is influenced by UVB radiation. Shedding of early-formed floral buds is usually the reason for any time delay of first flower observed in crop species. However, in many plant species, UVB radiation affects the size of flowers, anther number, pollen production, pollen germination, and pollen tube growth [170].

In general, the reproductive organs of most plants are highly protected. Sepals, petals, and ovary walls “screen” the reproductive organs from UVB. Pollen is



UV Effects on Living Organisms. Figure 21

Example of foliar symptoms (chlorotic patches) on bean leaves strongly enhanced by UV radiation exposure (Taken from [173])



UV Effects on Living Organisms. Figure 22

Comparison between leaves (from Patagonian *Jaborosa magellanica* Brisben) from nonirradiated plants (left) and leaves from irradiated plants (right) (Taken from [174])

therefore “at risk” when it falls on the stigma. Then pollen germination and rate of pollen tube growth are also affected by UVB radiation. A decrease in pollen tube growth by 10–25% may result. The fertilization process of sensitive plants is affected too, resulting in fewer seeds in these plants. However, the walls of the style and ovary may provide some protection once the pollen tube has penetrated the stigma.

Growth and Dry Matter Production Increases in length of organs (e.g., roots, leaf, or stem) or in weight are usual indicators of plant growth. For any given plant, several “quantities,” such as dry matter increase, increase in cell number, or increase in volume, are generally used to monitor growth [175]. For many plant species, studies in the field [176–179] or in controlled environment [180–182] have shown a decrease in growth of leaves and of stems caused by enhanced UVB radiation (Fig. 22). The observed growth reduction is explained by a reduction of the cell division rather than by a reduction in cell dimension. Reduction in plant height in connection with enhanced UVB radiation is explained by decreased growth hormone levels. On one hand, smaller and more compact canopies will better protect the plant by transmitting less solar radiation and UVB. On the other hand, the potential, or total, photosynthetic area, which is essential for growth, will be reduced [170]. Altogether, these factors will lead to a reduction in total dry matter and in biomass production. According to [170] 60% of crop species show a reduction in dry matter production after exposure to UVB radiation, 24% show no change, whereas only 8% show an increase in dry matter production.

Strategies for Protection Against UVB Radiation

During the process of evolution, plants developed repair and defense mechanisms by which UVB damages to the cell are limited and tolerance to solar UV-radiation is increased. Several UV-driven repair and defense mechanisms exist, which modify the optical characteristics of the leaves or other parts of the plant, or which work at the biochemical-molecular level [169, 170].

Repair Mechanisms

Cellular life forms usually possess repair enzymes that can recognize chemically modified bases, including those formed by UV radiation [170, 183]. In addition, a variety of biochemical mechanisms exist to restore the integrity of the genetic material after DNA damage, and thus to maintain its stability. These DNA repair mechanisms include photoreactivation, excision, and postreplication repair. During the process of photoreactivation, an enzyme (photolyase) responsible for the

splitting of pyrimidine cyclobutane dimers is involved. Excision repair consists in eliminating the damaged part of DNA by removing the bases in the damaged strand and then by synthesizing the gap. During the process of postreplication repair, the DNA damage is bypassed during DNA replication, and the resulting gaps are later filled in by using the sister duplex information. These kinds of repair mechanisms are observed in chloroplast and nuclear DNA [170, 183]. The ability to repair DNA varies a lot among the different plant species.

Defense Mechanisms

The first kind of defense mechanism is based on a better adaptation of the surface structure, physiology, and composition of the epidermal layer to attenuate the transmission of UVB through the epidermal layers, and eventually to better shield cells from UVB radiation [183]. The protective structures have the capacity to attenuate radiation damage by reflecting, absorbing, and scattering the incident flux. Among other mechanisms, hairs and wax coating have a prominent role [184]. UVB radiation leads to oxidative stress in plant systems, similarly to what is observed for abiotic and biotic stress. As a result, plants increase production of some chemicals such as flavonoids and antioxidant enzymes, which provide defense against UVB radiation [185–187]. Flavonoids, which are produced and mostly deposited in leaf hairs and in epidermal and mesophyll layers, are very efficient in mitigating the effects of UVB radiation. In this way, flavonoids may reduce damage to sensitive cell organs (e.g., DNA, chloroplasts, and mitochondria). Other compounds that could have the potential to protect against UV radiation effects are anthocyanins and carotenoids. They can protect the pollen grains, especially in flowers. They do not directly influence photosynthesis and other physiological processes, since they attenuate incoming radiation only in the UVB spectral range and not in the range of photosynthetic active radiation [170].

Selection or Genetic Improvement Plant species adapt and protect themselves from UVB in more or less adapted ways. The variable stimulation of protective and repair mechanisms may explain the difference among the various plant species. These variations

provide an opportunity for genetic improvement – either through traditional plant breeding techniques (crossing or selection) or through modern molecular biology techniques, such as plant transformation including some genetic modification. As a possible method, it may prove useful to identify tolerant species or cultivars by screening wide germplasm from various locations, and investigate the origin of UV tolerance [170].

Effect of UVB on Aquatic Plants and Macroalgae

In general, UVB radiation constitutes a significant stressor for macroalgae, aquatic mosses, liverworts, and aquatic flowering plants [130].

UVB affects macroalgae on the cellular, molecular, individual, and community levels [188]. UVA and UVB both affect photosynthesis and growth rate, and lead to accumulation of DNA damage. In addition, some studies report about the “impairment” of phototaxis used, e.g., by motile gametes of brown algae to accumulate at the water surface in order to increase the chance of finding a mating (do you mean: mating?) partner [189].

UVB sensitivity varies a lot among species. This results in vertical and in geographical “distributions” of the various species according to their sensitivity to UV. UV-tolerant species populate the tidal zone and are in general closer to the water surface. More sensitive species are found in deeper waters [190, 191]. Seasonal changes in UV and visible radiation also result in a succession of species over the year [192]. Macroalgae show an “ability” to adapt to UV. For instance, young specimens are more sensitive to UV than older specimens, and species collected shortly after the winter are found to be more sensitive than those harvested later in the year [193]. Most macroalgae have efficient repair mechanisms. In many macroalgae, efficient ROS (reactive oxygen species) (has ROS been defined earlier?) scavenging enzymes are found in addition to the DNA repair mechanisms.

Aquatic mosses and liverworts show a UVB-related inhibition of photosynthesis, growth, and pigmentation [194, 195].

For aquatic flowering plants, studies [130] show that the photosynthetic quantum yield dramatically decreases under unfiltered solar radiation, and that

removal of UVB or total UV leads to an increase in photosynthetic activity [196]. Experiments show also an efficient adaptation of sea grasses to solar UV [197].

UVB and Agricultural and Natural Terrestrial Plants

A meta-analysis comparing the overall reaction of woody and herbaceous plants under high supplemental UVB levels showed that the changes in “physiological” variables due to UVB observed in woody plants were significantly smaller than those observed in herbaceous plants [198].

Methodologies Used for the UV Experiments To investigate the effects of UV on plants, two different methodologies are followed:

1. Experiments in climate rooms and glasshouses
2. Experiments in the field

The experiments in climate rooms and glasshouses have the considerable advantage of a controlled environment. The drawbacks of this methodology are the unfavorable PAR to UVB ratio and differences between the PAR and UV spectra of the lamps and that of solar radiation. Results from indoor studies cannot therefore be easily extrapolated to outdoor conditions. Conversely, results from outdoor experiments are more realistic than those from indoor studies [199, 200]. Outdoor studies consist of experiments using some UV lamps to increase, or filters to decrease, the UV levels. In general, the increase or decrease in UV should be comparable to expected realistic changes in UV that may be simulated using a radiative transfer model.

UVB and Woody Plants Studies showed contradictory results regarding the effect of UVB on growth and photosynthesis in woody plants. Some investigations [201] demonstrated a sensitivity of woody plants to UVB irradiance, and also showed that these kinds of effects may be cumulative [202]. Other studies, however, found that even large increases in UVB radiation would not lead to any harmful effects on growth and photosynthesis in some woody plants [203, 204].

Radiative transfer in tree canopies is very complex: radiation incident on leaves depends on the orientation and inclination of the leaf and on the position of the leaf in the canopy. Whereas the upper foliage may be

fully exposed to UVB, the lower foliage is benefits from the shading by branches, leaves and twigs. This results in a very variable radiation environment with sunflecks and gaps. In various forests with closed canopies, as little as 1–2% of the incident UVB may be transmitted to the lower levels [205].

At the individual leaf level, surface reflectance and absorption by pigments are major determining factors for the plant's UVB sensitivity. In general the UVB surface reflectance of leaves is lower than 10% [205]. The UVB surface reflectance depends on leaf surface waxes and hairiness. Leaf hairs are one major protector against UVB radiation [206–208].

In general, the amount of UVB radiation reaching the photosynthetic mesophyll is relatively low in evergreen species and higher in deciduous species [209], due to a higher concentration of absorbing compounds in the epidermis of evergreen plants.

UVB and Herbaceous Plants Studies showed that herbaceous plants would suffer a negative effect in biomass production under elevated UVB radiation levels. Reductions of 7–14.6% have been found, compared to values under normal ambient UV levels. Elevated UVB leads to decrease in plant height and specific leaf area, as well as increases in herbaceous UVB absorbing compounds [198].

UVB and Field Crops Indoor or glasshouse experiments conducted with additional UVB lamps to increase UVB, or with filters to reduce UVB, usually show dramatic effects of UVB radiation. Due to the supplemental UVB sources, however, the tested UV levels are usually much higher than what would ever occur under natural outdoor conditions. Several studies also showed that it is much more complex to simulate and investigate the effects of changing UVB that would occur under natural conditions [172].

Many of the studies showed that enhanced UVB radiation levels could lead to visual symptoms consisting of chlorotic or necrotic patches on leaves exposed to UVB. UVB radiation alters the vegetative and reproductive morphology. Changes in leaf anatomy consisting of changes in thickness of epidermal, palisade, and mesophyll layers occur. After enhanced exposure to UVB, a decrease in chlorophyll content (10–70%), and an increase in UVB-absorbing

compounds (10–300%) are observed in many crops [172]. A decrease in photosynthesis (3–90%) is particularly observed at higher UV doses.

Yield and Yield Components The decrease in chlorophyll pigments and photosynthesis usually results in lower biomass and yield of crop plants. Some genotypes of crop species show a thicker leaf wax layer, a loss of chlorophyll, and an increase in phenolics (that allow a better UVB protection), ultimately resulting in changes in biomass and yield [172]. The effects of UVB on yield vary very much with crop species, however. In any case, most of the studies show that UVB increases lead to yield loss (Table 5).

Some species such as cowpea (*Vigna unguiculata*) [210, 211], millet (*Setaria italica*), and tobacco (*Nicotiana tabacum*) [212] show almost no yield reduction. Other species (e.g., pea [179], barley (*Hordeum vulgare*) [176, 179], mustard (*Brassica nigra*) [213], Black gram, Mung bean [214], and wheat [215]) show a strong reduction (Table 6). This yield loss consists of a reduced fruit grain number due to failure in fertilization, destruction of fruiting structures, and reduced fruit size due to decreased supply of assimilates to the growing sink (fruits). UVB also affects the yield quality. For instance, protein content and seed oil are reduced in soybeans that are exposed to enhanced UVB radiation.

Animals

Animals are divided here depending on the utilization of the species in the following four groups: pets (animals kept for companionship and enjoyment, or a household animal), farm animals (including horses for sport reasons), zoo animals, and finally fish and fishery-related animals.

A general finding is that animals show similar UV-induced changes of eye and skin as humans do. For the skin, two distinctive forms can be distinguished: malign melanoma and non-malign skin cancer, predominantly squamous cell carcinoma (SCC), and basal cell carcinoma (BCC).

Pets

UV Hazards The UV exposure of pets depends strongly on how these animals are kept. Cats and dogs

UV Effects on Living Organisms. Table 5 Selected results quantifying the influence of increased UV radiation on crop yield

Crop	Simulation of O ₃ depletion (%)	Experimental condition	Observed change in yield (%)	Reference
Barley	–	F	–17 to –31	Mazza et al. [176]
Black gram	–15	F	–63	Singh [214]
Corn	–20	F	–22 to –33	(Correia et al. [177])
Mung bean	–15	F	–76	Singh [214]
Soybean	–16	F, GH	–41	Teramura and Murali [216]
Wheat	–12, –20, –25	F	–43	Li et al. [215]
	15	F	15	Al-Oudat et al. [217]

F outdoor experiments, GH Greenhouse experiment

Source: Kakani et al. [172]

UV Effects on Living Organisms. Table 6 Sensitivity of selected crops to enhanced levels of UVB radiation. Results of experiments performed in controlled environments

Sensitive	Moderately sensitive	Relatively tolerant
Barley (<i>Hordeum vulgare</i>)	Common bean (<i>Phaseolus</i> spp.)	Corn (<i>Zea mays</i>)
Carrot (<i>Daucus carota</i>)	Lettuce (<i>Lactuca sativa</i>)	Cotton (<i>Gossypium hirsutum</i>)
Cucumber (<i>Cucumis sativus</i>)	Peanut (<i>Arachis hypogaea</i>)	Cowpea (<i>Vigna unguiculata</i>)
Mustard (<i>Brassica</i> spp.)	Pepper (<i>Piper nigrum</i>)	Clover (<i>Trifolium</i> spp.)
Oats (<i>Avena sativa</i>)	Petunia (<i>Petunia</i> spp.)	Millet (<i>Setaria italica</i>)
Pea (<i>Pisum sativum</i>)	Potato (<i>Solanum tuberosum</i>)	Radish (<i>Raphanus sativus</i>)
Soybean (<i>Glycine max</i>)	Rice (<i>Oryza sativa</i>)	Sunflower (<i>Helianthus annuus</i>)
Sweet corn (<i>Zea mays</i> var. <i>saccharata</i>)	Rye (<i>Secale cereale</i>)	Tobacco (<i>Nicotiana tabacum</i>)
Tomato (<i>Lycopersicon</i> spp.)	Sorghum (<i>Sorghum vulgare</i>)	Wheat (<i>Triticum aestivum</i>)

Source: Prasad et al. [170]

can stay outdoors for longer periods, whereas several other species that are kept as pets live predominantly indoor. For humans, it is well known that the incidence of melanoma and non-melanoma skin cancer depends on the UV exposure, which can be assessed by the geographic latitude, elevation, and/or annual global radiation (itself related to sunshine hours). For dogs, this correlation between environmental factors and the incidence of skin neoplasms could not be found in experiments conducted in Australia (Melbourne and

Queensland), USA, and UK [27, 218–227]. Two arguments could explain this discrepancy between humans and dogs: first, the fact that dogs stay predominantly indoors, so that the influence of environmental factors is reduced, and second, the different distribution of breeds in these regions. For farm animals, it is well known that some breeds are not adapted to areas with high UV exposure due to lack of pigmentation (e.g., Herford cattle).

Teifke and Lohr [228] analyzed 106 SSC immunohistochemically for overexpression of p53,

which is an indicator of a UV-related incidence. In 9 of 11 (82%) feline SCCs of the ear and in 7 of 14 (50%) feline SCCs of other locations, p53 immunoreactivity was detected. For dogs, 7 of 25 (30%) cutaneous SCCs gave a positive reaction. This is a good indicator that UV exposure is an important external factor for SCC.

The clinical picture of eye cancer, such as cutaneous SCCs in the close vicinity of the eye, could be observed in guinea pigs, hamsters, rats, and mice, which are used as pets as well. Pigmented rats and mice showed a significant reduction in the incidence of development of such cancerous eyes compared with nonpigmented animals [229].

Dogs A study on humans estimated that 0.25–1% of actinic keratosis (AK), which is mainly caused by UV exposure, progress toward SCC. Similar data do not exist for cats or dogs. Nevertheless, it can be observed that AK is common in the vicinity of SCC. In dogs, they are found primarily on the central abdomen; ventral flanks; and medial thighs of short-coated, white-haired, or piebald breeds. Dalmatians, Pit Bull Terriers, Beagles, Basset Hounds, and other dogs with similar coat characteristics have an increased incidence of AK, as well as other solar-induced neoplasms [230]. The occurrence of AK on the nose is often called “collie nose” (nasal solar dermatitis), even though it is not restricted to this breed [231]. As possible therapy, the application of sun screens with a high sun protection factor, or the tattooing of the nonpigmented areas of the nose, is recommended.

A genetically homogeneous group of 991 Beagles was housed in gravel-based, outdoor pens with dog-houses in a high-altitude, high-sunshine environment, in Colorado (40.5°N, USA). Solar dermatosis was restricted to the sparsely haired, nonpigmented abdominal skin. Solar dermatosis was diagnosed in 363 of the 991 dogs, representing an incidence of 36.6%. There were 175 cases of hemangiomas, hemangiosarcomas, or SCC of the skin among the 991 dogs. Of these, 129 tumors occurred in dogs with, and only 46 in dogs without, solar dermatosis [232].

Nielsen and Cole [233] found that there is no preferred location for SCC, contrarily to humans, whose preferred sites are the most UV-exposed areas of the body.

The influence of age is mentioned by Nesbitt [234], who observed the occurrence of SCC mainly for dogs

older than 5 years. The cumulative incidence for such lesions for Beagles was observed with 47% at the age of 12 years [235]. For a genetically homogeneous group of 991 Beagles, Nikula et al. [232] investigated the role of age on the cumulative incidence of hemangiomas, hemangiosarcomas, and SCC. For SCC, the slope gets steeper for an age of 7 years (Fig. 23).

A breed disposition for the incidence of SCC was observed for dogs with insufficient pigmentation on the lips and nose, as well as on toes and ventral abdomen. Worse affected are: White German Shepherd, White Australian Shepherd, and Welsh Corgi [234].

Cats In the case of cats from sunny places, SCC could be found on the head. Risk factors are UV radiation, nonpigmented skin, and thin hair [236]. Actinic keratoses occur most often on the pinnae, nose, and eyelids of white-faced cats [230]. The influence of pigmentation was shown by Dorn et al. [237] for white-faced, blue-eyed cats. The prevalence of SCC in white cats is 11-fold higher than in nonwhite cats (Fig. 24).

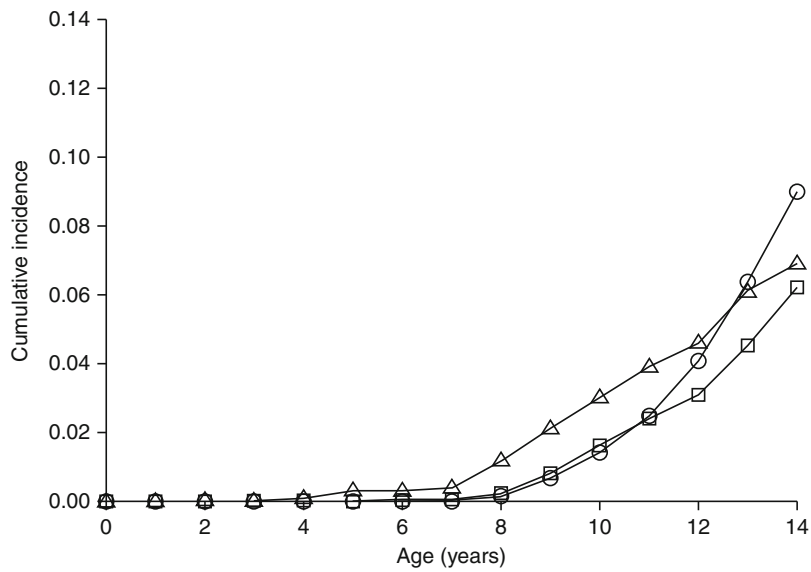
In cats, melanomas could be hardly observed. After Jörger [236], no race-specific or sex-specific predispositions are known.

Vitamin D in Pets

Cats and Dogs Vitamin D is synthesized by UV in the skin of omnivores (rats, pigs, humans) and herbivores (horses, sheep, cattle). On the contrary, carnivores are solely dependent on oral intake to meet their vitamin D requirement [220–222].

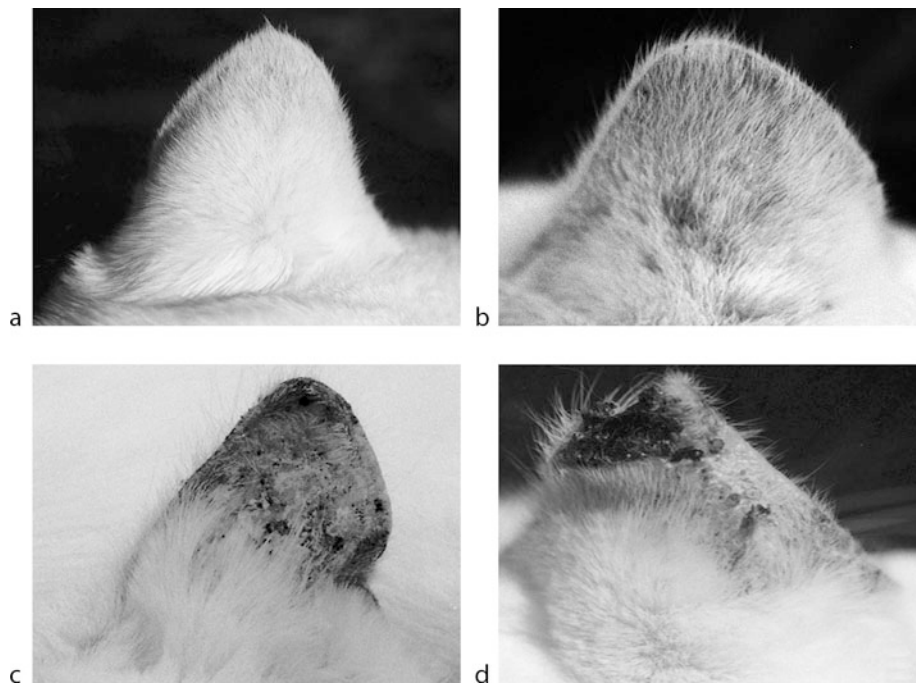
How et al. [223] have shown that there is no photosynthesis of vitamin D₃ in cat skin. Although not strictly carnivores, dogs lack in the ability to photosynthesize vitamin D₃ [223]. It could be shown that hypovitaminosis D occurs in dogs irradiated with UV but with no vitamin D in their food [220].

Both natural food and commercially available complete dog food contain sufficient vitamin D₃ to fulfill a dog's requirement. Therefore, in domesticated dogs and cats, rickets is only seen under extreme circumstances, such as a strict vegetarian diet, biliary atresia, and inborn errors of vitamin D₃ metabolism [226]. Hypervitaminosis D and coupled hypercalcemia are rarely reported [225].



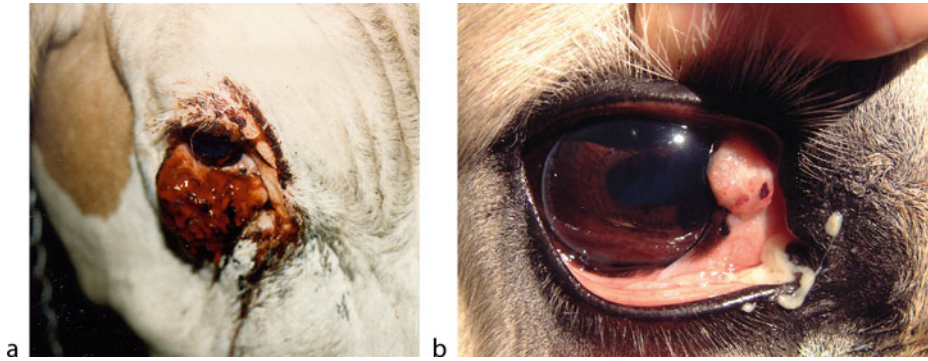
UV Effects on Living Organisms. Figure 23

Cumulative incidence of hemangiomas (○), hemangiosarcomas (□), and SCC (Δ) for Beagle dogs (Source: Nikula et al. [232])



UV Effects on Living Organisms. Figure 24

Demonstration of progressive stage of photo damage shown on the ear of cats. (a) Normal, (b) initial photo damage, with erythema and scaling, (c) advanced photo damage, with erythema, scaling, erosion, ulceration, and plaques, and (d) advanced photo damage and SCC (Source: Almeida et al. [238])



UV Effects on Living Organisms. Figure 25

Cancer eye (ocular squamous cell carcinoma) around a sparsely pigmented skin area (a) cattle (Source: [9]) and (b) horse (Source: [241])

Rabbits In rabbits, there is a clear difference in vitamin D concentration between animals housed in hatches and animals kept in free range. A strong seasonal variation in plasma vitamin D₃ concentration has been reported in rabbits [219]. For animals that had the opportunity to bask in the sun, the highest concentration was found between May and September. It takes about 5 months for the vitamin D reserve to become depleted after exposure has stopped, in conjunction with a vitamin D deficient diet [224]. Vitamin D content of forage and vegetables being negligible, only high-quality hay could contribute to the vitamin D level in rabbits.

Rats The fact that UV radiation cures rickets in rats was reported by Sonne and Rekling [27]. Beside a dose-response relationship, they were also able to derive an action spectrum. This action spectrum is quite similar to that for humans, with high efficiency in the UVB. How et al. [222] showed directly that after UV exposure an increase in vitamin D₃ in rats could be observed.

Farm Animals

UV Hazards Beside the known short-term effects like erythema, several long-term effects can be detected, such as elastosis and skin cancer. Farm animals of several species – cattle [9, 239], goats, sheep [240], and horses [241] – can develop skin cancer, mainly squamous cell carcinoma (SCC), in sparsely haired and sparsely pigmented areas of the skin.

Several farm animal species can also develop the so-called cancer eye, which is a squamous epithelial

carcinoma near the eye. This disease typically occurs because the eyelid is hairless, and is dependent on both pigmentation and UV exposure. This cancer eye (Fig. 25) can be observed predominantly in Herford and Simmentaler cattle, Ayreshire, Shorthorn, black-pied cattle, Indian water buffalo, and their cross breedings [9, 239], but also in horses [241].

Teifke and Lohr [228] analyzed 106 SCC immunohistochemically for overexpression of p53, which is an indicator of a UV-related incidence. All of six (100%) equine ocular SCCs and seven of nine (78%) SCCs of the equine penis or vulva gave positive reactions. Therefore, UV radiation can be assumed as a major etiological factor for SCC.

Cattle Heeney and Valli [242] reviewed several epidemiological studies about bovine ocular SCC (cancer eye). Etiological factors were UV radiation, circumocular apigmentation, and viruses. Concerning pigmentation, Hereford cattle are commonly affected. The incidence of ocular lesions was found between 8.1–45.7% [243], 35–58% [244], 21% [245], and 25–35% [242]. For Holstein cattle, a much lower incidence was observed, <1% [242], as a result of a denser circumocular pigmentation.

In a study of Gharagozlou et al. [246], 32 cases of ocular neoplasms were diagnosed. The affected animals were female (100%), adult and more than 50% of them aged more than 5 years. In most cases (70%), the lesions were located in the nictitating membrane and palpebral conjunctiva. Intraocular invasion was noted in seven cases (22%). Microscopically, in 12 cases out of

32 (38%) the tumors were noninvasive SCC or carcinoma in situ; 18 cases (56%) were invasive SCC; a single case (3%) was lymphosarcoma, while a further single case (3%) was malignant hemangioendothelioma.

Morris [247] investigated the heritability of predictors that are closely related to skin cancer. The heritability of the eyelid pigmentation was estimated at 0.64–0.83. The Hereford cattle also appear to be highly sensitive to eye cancer. Estimates indicated a moderate heritability: 0.17–0.29 for eye cancer, 0.10 ± 0.08 for the incidence, and 0.30 ± 0.09 for the number of tumors.

The influence of UV exposure on the incidence of cancer eye can be related to geographic latitude, duration of sunshine, and altitude. Anderson and Skinner [248] classified 5,000 Hereford cattle into groups, describing the UV exposure based on latitude, sunshine duration, and altitude (Table 7). The stratification of the data showed that for all three parameters, the incidence of skin cancer stays proportional to the UV exposure.

Anderson and Badzioch [245] performed epidemiological investigations on 34 Hereford herds in North America (Table 8). For some parameters that are correlated with UV intensity (sunshine duration, elevation, cloudiness index, and global radiation) a relationship with ocular SCC was found. The odds ratios for these parameters, as a measure for the risk to

develop a SCC, are summarized in Table 8. Because of the similar geographical latitudes of the statistical investigated territory, these results can be applied to midlatitude countries and southern Europe as well.

In Zimbabwe, ocular SCC was frequently observed in Simmental cattle, and exposure to intense solar radiation has been proposed as the cause, especially when cattle are kept at high altitude (1,500 masl) in a sunny and warm climate [249]. No cases were observed for fully pigmented cattle breeds.

The influence of the age t (in years) of Hereford cattle on the incidence I (in %) is depicted in Fig. 26, showing an exponential growth according to $I = 0.519 \exp(0.193 t)$ (with an adjusted coefficient of determination $r^2 = 0.541$), based on the data from Woodward and Knapp [250] and Russel et al. [244]. The half value period gives 3.6, which means that the incidence I doubles every 3.6 years. The first data set was collected in Montana at a latitude of about 46°N , and the second one in Colorado at a latitude of about 37°N . Anderson and Badzioch [245] found a 2% increase of risk per month of age, compared to 1.6% per month for the exponential model in Fig. 26. For New Mexico, Blackwell et al. [251] found an incidence for ocular SSC which was 2.5 times larger than in the two previous data sets. This could be caused by different climatic conditions, combined with a lower geographic latitude of about 35°N . Therefore, these data

UV Effects on Living Organisms. Table 7 Relationship between solar UV exposure and the incidence of cancer eye in Hereford cattle

Criterion	UV exposure	# animals	Incidence	Arithmetic mean within each level		
				Latitude ($^\circ$)	Sunshine (h)	Altitude (masl)
Latitude ($^\circ$)	Low	3,445	3.6 ± 0.3	46.4	2,593	809
	Medium	361	7.8 ± 1.4	43.4	2,840	1,143
	High	1,154	9.2 ± 0.8	33.6	3,150	776
Sunshine duration (h)	Low	3,445	3.6 ± 0.3	46.4	2,593	809
	Medium	823	5.6 ± 0.8	38.1	2,800	691
	High	692	11.7 ± 1.2	33.4	3,400	1,068
Altitude (masl)	Low	670	4.8 ± 0.8	36.1	2,880	357
	Medium	3,909	5.1 ± 0.1	44.4	2,718	846
	High	381	9.7 ± 1.5	43.9	2,704	1,433

Source: Anderson and Skinner [248]

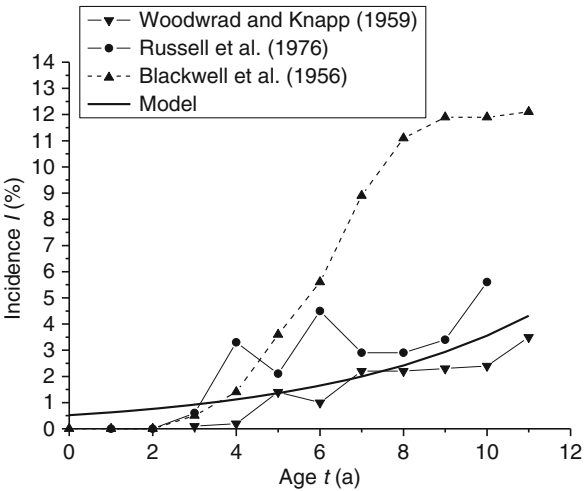
UV Effects on Living Organisms. Table 8 Relationship between solar UV exposure and the incidence of cancer eye in Hereford cattle in North America (After Anderson and Badzioch [245]) (Global radiation is expressed in ly (1 ly = 4.184 J/cm²), and cloudiness index as the global horizontal irradiation normalized with its extraterrestrial equivalent)

UV exposure described by	No of animals	Affected (%)	Odds ratio
Sunshine duration (h)			
1,800–2,299	273	13.6	1.0
2,300–2,599	335	16.1	1.1
2,600–2,899	1,158	21.2	1.8
2,900–3,499	1,009	25.7	2.5
Global radiation (ly)			
200–299	283	17.3	1.0
300–399	1,455	19.1	2.1
400–499	1,037	25.8	3.1
Cloudiness index (–)			
0.400–0.499	1,067	13.6	1.0
0.500–0.599	690	27.2	3.4
0.600–0.699	1,018	25.7	2.9
Altitude (ft)			
0–1,999	1,308	13.7	1.0
2,000–2,999	564	30.3	2.8
3,000–3,999	381	33.6	4.6
4,000–9,200	522	22.4	3.2

were not included into the regression model for the incidence (Fig. 26).

For cattle, SCC could be observed also on the vulva [232]. Wettmuny [252] found that pigmentation is a main criterion for the incidence of vulva carcinoma in Ayreshire cattle in Ceylon (8°N; 1,800 masl). The same relationship was obtained by Hünernmund [253] for Kenya (1°S, 2,200 masl). The incidence of SCC on nonpigmented areas of the vulva reached between 1% in Israel (30°N) [254] and about 10% in Ceylon [255].

Like the skin, the eye is also sensitive to UV exposure. In an epidemiological survey about infectious keratoconjunctivitis in Australian cattle, Slatter et al. [239] derived a relationship between season and



UV Effects on Living Organisms. Figure 26 Influence of age (years) of Hereford cattle on the incidence of cancer eye $I = 0.519 \exp(0.193 t)$ (%) (Data for the regression model from [244, 250])

frequency of disease. Similarly, Kopecky et al. [256] established that solar radiation was an amplifying factor for this disease.

Small Ungulates (Sheep and Goat) For sheep and goat, SCC could be observed on the vulva, eyelid, ear, and other parts poorly covered by wool [232].

The clinical and pathological characteristics of the outbreak of SCC observed in that study suggest that the breed of sheep, its lack of skin pigmentation, the farm's southern latitude (where UV is comparatively stronger than at the same latitude in the northern hemisphere), and factors related to the type of farming played a decisive role in the development of SCC in these sheep [240]. For Marino sheep, Forrest and Fleet [257] could show the lack of tanning capability as well as a distinct erythematous response to UV radiation. Due to a long-term UV exposure of over 28 days, the thickening of the epidermis from 23 to 120 μm was observed.

Horse For horses, a histological examination of 21 suspected neoplasms confirmed the presence of SSC (76% of cases). The remaining cases were diagnosed as lymphoid hyperplasia (14%), mast cell tumor (5%), and sebaceous gland adenocarcinoma (5%) [241]. Beside SCC, melanoma is also a common equine tumor [258]. Approximately 80% of gray horses over

15 years of age show melanotic growths. The neoplasms are generally found at the perineum, vulva, and undersurface of the root of the tail, but also on the male genitalia, neck, and ears. Melanomas remain benign for 10–20 years on the skin; however, the disease is rapidly fatal once vital organs are involved.

Genetic disposition of horses for ocular SCC were found by Dugan et al. [259], who analyzed the databases of 14 veterinarians in the USA. It was found that breed and hair color had a significant effect on the prevalence of ocular SCC. The increasing prevalence with increasing age can be calculated by the odds ratio (OR). The influence of UV radiation on the prevalence of ocular SCC was found as a function of longitude (do you mean: latitude?), altitude, and annual solar exposure [259], in good agreement with epidemiological studies for other species (e.g., cattle [245]).

The adjusted OR results summarized in Table 9 provide the relative prevalence $P_{\text{rel}}(x, y)$ describing the change of prevalence of an older horse (age x , in years) in relation to a young horse (age y , in years) according to $P_{\text{rel}}(x, y) = OR^{x-y}$. The age effect on OR is 1.1 ($p < 0.001$). This means, for instance, that a 12-year-old horse has a 2.6-fold higher risk to develop ocular SCC than a 2-year-old horse (of the same breed and hair color) based on $P_{\text{rel}} = OR^{12-2} = 1.1^{10}$.

UV Effects on Living Organisms. Table 9 Adjusted odds ratio (OR) of variables that are correlated with the occurrence of ocular SCC in horses (After Dugan et al. [259]). Variables with a nonsignificant OR ($p < 0.05$) are not shown here

Variable		OR	p-value
Breed	Mixed breed	3.1	<0.001
	Paint, pinto	4.5	<0.001
	Apaloosa	7.6	<0.001
	Draft breed (Belgium, Clydesdale, Shire)	21.6	<0.001
Hair color	Chestnut, sorrel	3.8	<0.001
	Buckskin	4.4	<0.01
	Red/white or strawberry/white	4.7	<0.001
	Gray	6.7	<0.001
	Creamello, polomino	13.7	<0.001
	White	26.7	<0.001

Photosensitivity Photosensitivity defines an abnormal reaction of the skin to visible light and UV radiation. Basically, two forms of sensitivity can be distinguished: Phototoxic reactions and photo-allergic reactions. The former occur after absorption of radiation by molecules, and the membrane and DNA damages that follow. The reaction of the skin is restricted to exposed areas. Photo-allergic reactions are a more seldom phenomenon, where the agent evokes immunologic reactions. An observed photo-allergic reaction can persist many years, even when the causing noxa has already been eliminated.

Photosensitive substances can be found in either drugs or feed. For instance, buckwheat (*Fagopyrum esculentum*) is known to cause photosensitivity.

This disease can be observed mainly in sheep and swine, rarely in cattle and goat, and hardly in horse. Saint-John's-wort (*Hypericum perforatum*) causes a disease in sheep, which manifests as a depigmentation by hypericin. Other sensitizers exist, such as furanocoumarin, which can be found in apiaceae or umbelliferae (like hogweed [*Heracleum mantegazzianum*], chervil [*Anthriscus cerefolium*], angelica [*Angelica archangelica*]) and yarrow (*Achillea millefolium*).

The application of veterinary medicines and medicated feed may expose agricultural workers to antimicrobial drugs, tranquilizers, and other chemicals with phototoxic and photoallergic side effects, like phenothiazin, acriflaviniumchlorid, tetrazyklin, mansonil, or sulphonamide. Olaquinox, a growth promoter for swine, is known to cause photosensitization in farmers, probably due to a photoallergic mechanism. Moreover, olaquinox was found to be phototoxic in animal experiments [260–263].

Vitamin D Vitamin D₃ is synthesized by UV radiation in the skin of omnivores (pigs, humans) and herbivores (horses, cows, sheep, goats). This synthesis works in fur-bearing animals even in the presence of hair.

For animals, both types of vitamin D – namely, cholecalciferol (vitamin D₃) and ergocalciferol (vitamin D₂) – are of importance. Whereas vitamin D₃ is produced in the skin, vitamin D₂ is ingested with feed.

Vitamin D is negligibly low in forage and silage, as well as in corns, roots, seeds, and grains. The only exception is silage made of maturity corn. Contrarily, hay contains a significant amount of vitamin D.

For animals in agricultural use, the antirachitic properties of vitamin D are brought by vitamin D₂ supplementation to food [8]. From animal welfare reasons outdoor areas for farm animals are also quite common. In such a case, supplementation is not necessary [264]. The free-range practice can also be a problem, because most swine have a very low minimal erythema dose (in MED) due to a lack of photoprotection by hair coat or low pigmentation. Exceptions are found in older races, like Mangaliza pigs, which are well adapted to solar UV radiation.

In calves, exposition to UV was shown to enhance weight gain [265] compared to nonexposed calves.

Cattle can receive vitamin D from two sources: either UV radiation or plants exposed to the sun. Vitamin D₂ is produced by UV radiation by ergosterol in forage. However, common forage (grass, grass silage) is often poor in vitamin D [266]. This explains why, in cattle, vitamin D₃ is the major form in cattle blood plasma, whereas vitamin D₂ is minor [267]. When cattle are brought in the sun or specially exposed to UV, vitamin D₃ concentrations increase within the first 5 weeks of exposure up to a plateau level [268]. In pasture animals, the annual cycle of vitamin D₃ is correlated with the annual course of solar UV exposure [269, 270]. A plateau is held from June to August. In animals that have no longer access to UV radiation, the plasma concentration in vitamin D₃ decreases rapidly, whereas the concentrations of calcium and magnesium remain much longer, and even increase for a certain period after UV exposure has stopped [268, 269]. The rather rapid decrease in vitamin D concentration may be caused by tits release into the milk [269]. Milk contains only low vitamin D concentrations – even in summer when cows are at the pasture [269]. The vitamin D content of milk can be enhanced by a factor of 100 just by irradiating it with UV.

In sheep, vitamin D is gained from both dietary sources and photosynthesis in the skin exposed to UV. In the grazing sheep, D₃ plays a dominant role in the vitamin D status [270]. When sheep are kept indoors, they need a diet that delivers vitamin D; otherwise, vitamin D supplementation has to be given. In sheep,

D₂ is the dominant form of vitamin D in the blood plasma [264]. However, the presence of vitamin D₃ increases with UV exposure [271, 272], although there seems to be no linear relationship between vitamin D₃ and UV exposure. Experiments have shown that sheep easily reach their optimum vitamin D₃ level through UV exposure. Plasma D₃ and 25-OH D₃ are still detectable 7 days after the UV exposure has stopped (whereas it lasts 14 days in the case of D₂ from diet) increasing for the next 7 weeks up to a plateau. The UV caused vitamin D concentration plateau was two times higher than the plateau caused by vitamin D supplemented diets. After treatment (UV or diet), the vitamin D level stays stable twice as long as that from diet. It is concluded that UV is faster and more effective than diet to improve the vitamin D level.

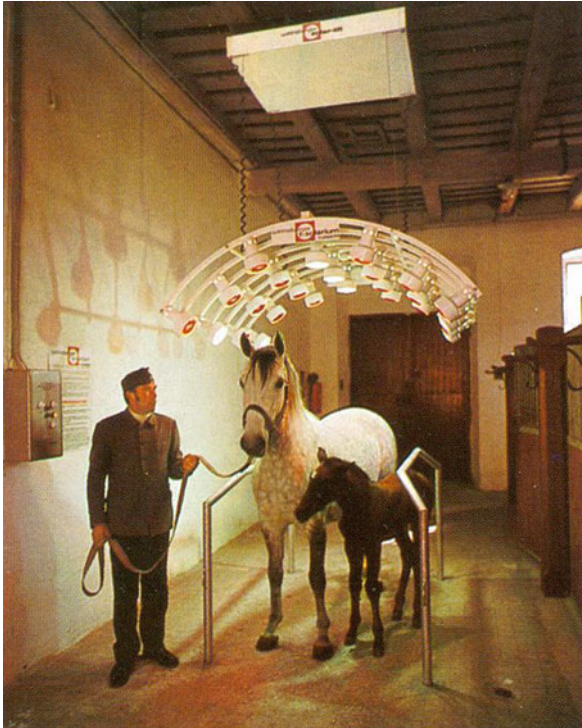
Sheep kept outdoors show strong seasonal variations in plasma 25-OH D₃ concentration [273]. In winter, its level depends strongly on geographical latitude.

Wool has an important influence on the vitamin D level. As shown by Zintzen and Boyazoglou [274], the concentration of vitamin D in shorn sheep was two to three times greater than in unshorn animals. With increasing Fleece length, the vitamin D level decreases.

For goats and pigs, studies on vitamin D with respect to UV radiation are very rare.

Horses exposed to 4–6 h of solar radiation (including cloudy days) can synthesize sufficient quantities of vitamin D [275]. Solaria for horses (Fig. 27) have been recommended for animals that were held predominantly in stables, to compensate for the lack of solar UV exposure. Such radiation devices use lamps for visible and infrared radiation, as well as UV lamps, to simulate the entire spectrum of the Sun. From this exposure to simulated solar radiation, several positive physiological effects are expected. The spatial distribution of UV radiation over the entire horse body, as received either from the sun or from solaria, has been investigated by Keck et al. [55], Fig. 28.

This study showed that the UV skin dose from outdoor exposure had a totally different spatial distribution than that resulting from exposure beneath a solarium, as depicted in Fig. 28. The UV dose distribution was measured by 17 polysulphon dosimeters that were fixed with mastic on the hair coat of the horses. The dose distribution beneath the solarium



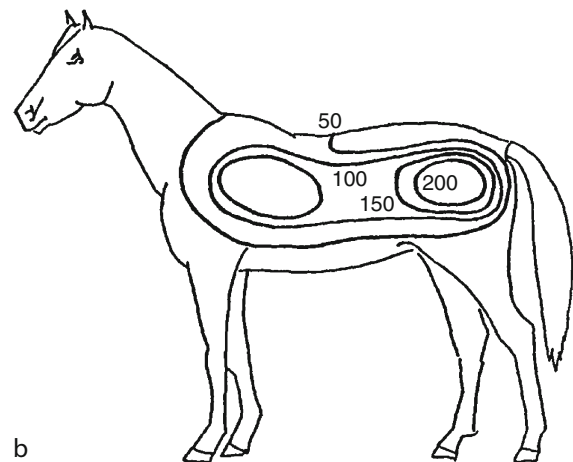
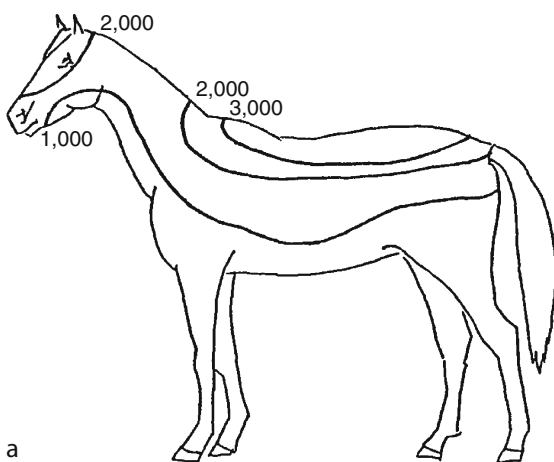
UV Effects on Living Organisms. Figure 27
Application of a solarium for horses in the Spanish Riding School of Vienna (Source: Weinsberger SW GmbH, Germany)

depends on the geometry of the UV lamps (Fig. 27), whereas the solar radiation field is much more homogeneous due to the presence of hemispherical diffuse radiation from the sky. Consequently, in outdoor conditions, hairless or little haired regions of the horse's skin receive a considerable portion (up to 60%) of radiation compared to a horizontal receiving surface. This means that UV effects can be assumed especially in these body regions.

The local maximum of the UV dose was $\approx 3,500 \text{ J/m}^2$ for outdoor exposure and $\approx 400 \text{ J/m}^2$ for solarium exposure. Considering a UV transmission of the hair coat of horses of less than 2% [55] the effective skin dose is about 70 J/m^2 for outdoor exposure and less than 9 J/m^2 for the solarium. Based on the Grotthus-Draper law, there is only little evidence that the (low) skin dose of horses beneath a solarium can contribute to positive UV effects like vitamin D synthesis. Longer exposure times, or more powerful UV lamps, would have to be used.

Horses exhibit a strong diurnal variation (50%) in vitamin D₃ serum level [276, 277]. This phenomenon seems to be restricted to horses.

Chicken can gain vitamin D from feed and UV exposure. Early studies [278–281] on poultry indicated that 10–30 min of daily sun exposure could prevent all



UV Effects on Living Organisms. Figure 28
Spatial distribution of the biologically effective UV dose outside of the hair cover of horses (J/m^2) (a) typical 1-day outdoor exposure, (b) solarium with a recommended exposure time of 22 min (Source: Keck et al. [55])

signs of vitamin D deficiency in growing chicken. It has also been known for a long time that a 15-min exposure to a quartz mercury vapor arc lamp produced similar effects [281].

Because of the dense plumage covering the chicken body, photosynthesis occurs in the featherless skin of legs and feet. The chicken's skin of the legs and feet contains eight times as much 7-dehydrocholesterol as their body skin [282]. Bernard et al. [283] pointed out that UV exposure (from lamps) and vitamin supplementation of food had similar efficiency. Higher hardness of egg shells can also be observed in chicken with sufficient vitamin D supply [8].

A known problem with chicken is that viable eggs produced by apparently healthy adults may incubate full term but then fail to hatch. The fully developed, dead embryos appear normal but have poorly mineralized skeletons. This problem can be successfully corrected by providing UV to the adult female prior to oviposition [284–287].

Fish and Fisheries

UV Hazards Fish and fisheries are part of the aquatic ecosystem, which produces more than 50% of the total biomass of our planet. The primary producers (phytoplankton) in freshwater and marine ecosystems constitute the basis of the food webs, providing energy for the primary (e.g., zooplankton) and secondary (e.g., fish) consumers and are thus important contributors to the production of staple diet. Humans use about 8% of the productivity of the entire oceans, but this amounts to 35% of the productivity of temperate continental shelf systems [130].

Phytoplankton and bacterioplankton are primary producers, since they use solar energy for photosynthetic conversion of CO₂ and nutrients into carbohydrates (in the so-called photic or euphotic zone). They account for about 95% of the primary productivity of oceans, and about half of all primary productivity on earth. Phytoplankton occurs predominantly in cooler, midlatitude zones with sufficient nutrients, especially nitrogen. The two major primary producers are the diatoms, which are located in temperate and polar oceans. They contribute about 60% of the primary productivity of oceans, with a typical size of about 30 µm. In contrast, the coccolithophores, with a typical

size of 5–10 µm, dominate in regions of moderate turbulence at midlatitudes in late spring, as well as in subpolar and equatorial regions. They contribute about 15% to the primary productivity of oceans.

Zooplankton is the primary consumer of phytoplankton. They range in size from single-celled organisms to larger multicelled organisms. Small zooplankton is eaten by larger zooplankton. Zooplankton include ciliates (single-celled animals), copepods, shrimp, and larval forms of barnacles, mollusks, fish, and jellyfish.

The UV sensitivity of primary producers (phytoplankton) and consumers (zooplankton) in the food chain is an important predictor of biomass production. The UV protection capability of plankton is inversely related to its size. Bacteria and nano- or pico-plankton are too small to effectively protect themselves against UV radiation by absorption. There is an upper limit for the concentration of absorbing substances due to osmotic restrictions [288]. DNA damage correlates strongly with the penetration of UVB radiation into the water column. As long as the repair mechanisms keep up with the UV-caused damage, the plankton population is not threatened. However, when cyclobutane pyrimidine dimers accumulate under high-UV exposure, the population decreases.

The UV sensitivity of aquatic plankton (bacterioplankton, phytoplankton, and zooplankton) can be expressed by the radiation amplification factor, RAF, which is defined as the relative increase of UV-caused damage for a relative change in total ozone (Table 10).

By using a model that describes the UV-induced DNA damage in oceanic bacterioplankton, the sensitivity on varying ozone thickness (which conditions the UV exposure at the sea surface), dissolved organic matter concentration, chlorophyll concentration, wind speed, and mixed layer depth was investigated. From the model, the total amplification factor (TAF; a relative measure of the increase in UV damage associated with a decrease in ozone thickness) for net DNA damage in the euphotic zone is 1.7, as compared to 2.1–2.2 for UV radiation weighted for damage to DNA at the surface [291]. The link between producers and consumers in the food web has been successfully shown by Kouwenberg and Lantoiné [293] for trophic plankton interactions. Diatom (*Skeletonema costatum*) was exposed to UV radiation in doses of 10% higher

UV Effects on Living Organisms. Table 10 Radiation amplification factor (RAF) for aquatic species

Species	RAF
Phytoplankton motility (<i>Evglena gracilis</i>)	1.5–1.9
Phytoplankton photosynthesis (<i>Phaeodactylum</i> sp.)	0.2–0.3
Phytoplankton photosynthesis (<i>Prorocentrum micans</i>)	0.3–0.4
Phytoplankton photosynthesis, in Antarctic community	0.8
Phytoplankton photosynthesis (<i>Nodularia spumtgena</i> cyanobacteria)	0.2
Bacterioplankton DNA damage (euphotic zone)	1.7
Bacterioplankton DNA damage (surface water)	2.1–2.2
Green alga <i>Prasiola stipitata</i>	0.7
Dinoflagellate <i>Gyrodinium dorsum</i>	0.4
Cyanobacterium <i>Anabaena</i> sp	1.0
Corals photosynthesis	0.21

Source: References [289–292]

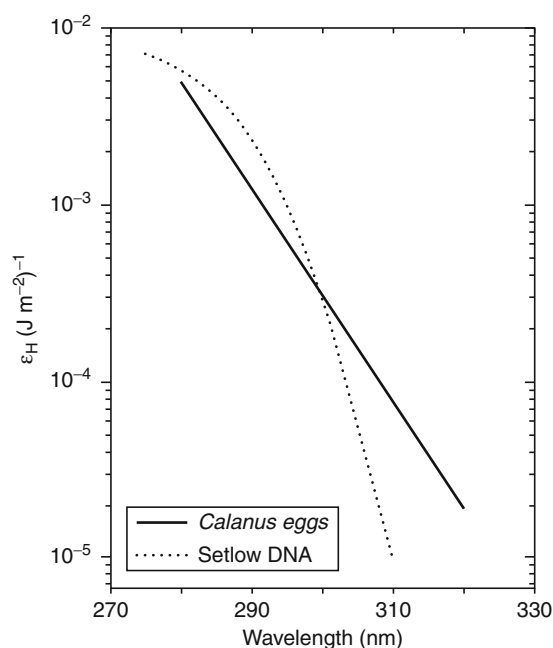
than the natural UV exposure. UVB-exposed algae showed modifications in cell structure, volume increases, and delay in cell division. These UVB-stressed *S. costatum* cultures were used as food for wild copepods (*Calanus helgolandicus*), which showed an indirect UVB effect on their reproductive output.

The sensitivity of copepoda *Calanus finmarchicus* to UV is an important factor conditioning the planktonic food chain of the St. Lawrence Gulf, Canada [294]. The biological weighting function of the mortality of copepods eggs is similar to that of DNA damage (Fig. 29).

Based on the spectral UV sensitivity of copepods eggs, their mortality corresponding to solar UV irradiance at noon was calculated by an exponential law

$$N = N_0 \exp(-E_{\text{biol}} t) \quad (15)$$

where the egg number of live eggs is N_0 , and N is the egg number after exposure of the eggs to a biologically effective irradiance E_{biol} during a time t . The weighting of the irradiance was done by a weighting function that was derived from Fig. 29. For an egg mortality of 50% at the sea surface, a 2.5 h exposure time was assumed.



UV Effects on Living Organisms. Figure 29

Action spectra of the copepod egg mortality and of the DNA damage [81] (Source: Kouwenberg et al. [294])

At a depth of 50 cm the exposure time was rather 4.6 h, due to the UV absorption of the water column [294].

To investigate the natural protection of Antarctic marine organisms against solar UV exposure, 57 species (1 fish, 48 invertebrates, and 8 algae) were collected during austral spring (Anvers Island, Antarctic Peninsula), and were analyzed for the presence of mycosporine-like amino acids – compounds that absorb UV radiation and may provide shielding from UV radiation. Nearly 90% of the 57 examined species contained mycosporine-like amino acids [311].

To assess the ecological UV sensitivity of plankton and aquatic animals, the vertical distribution of the species (e.g., Uitz et al. [312], Kesler et al. [313]) and the spectral absorption of the water column have to be taken into account [85, 314, 315].

For fish, several experiments were done to investigate their sensitivity to solar UV radiation. In addition to the mortality of eggs, skin lesions and eye damages were observed and summarized (Table 11). Rodgers [305] referred to the UV damage of the skin as the “summer syndrome.” Photobiologic responses in fish are also referred to as “sunburn” [308].

UV Effects on Living Organisms. Table 11 Biological UV effects in fish

Biological effects	Fish species	Author
Skin lesions, damage on eggs and larvae; damage of the brain, eyes, disorders in growth and development	Anchovies (<i>Engraulis mordax</i>)	Hunter et al. [295]
	Pacific mackerels (<i>Scomber japonicus</i>)	
Dose-response functions for mortality of eggs	Sockeye salmon (<i>Oncorhynchus nerka</i>)	Bell and Hoar [296]
Dose-response functions for mortality	Galaxiids (<i>Galaxus maculatus</i>)	cit. by Zagarese et al. [297]
	<i>Odontheistes hatcheri</i>	
Dose-response functions for mortality of eggs	Atlantic cod (<i>Gadus morhua</i>)	Kouwenberg et al. [298]
Skin lesions, sunburn	Chinook salmon	Brocklebank and Armstrong [299]
	Arctic char	
Skin lesions, sunburn	Hammerhead shark (<i>Sphyrna lewini</i>)	Lowe and Goodman-Lowe [300]
Skin lesions, sunburn	Juvenile rainbow trout (<i>Salmo gairdneri</i>)	Dunbar [301], Bullock and Coutts [302]
Skin lesions, sunburn	Juvenile paddles fish (<i>Polyodon spathula</i>)	cit. by Zagarese et al. [297]
Skin lesions, sunburn	Juvenile plaice (<i>Pleuronectes platessa</i>)	cit. by Zagarese et al. [297]
Skin lesions, sunburn	Atlantic salmon (<i>Salmo salar</i>)	McArdle and Bullock [303], Kaweewat and Hofer [304], and Rodger [305]
Skin lesions, sunburn	Koi carp (<i>Cyprinus carpio</i>)	Bullock et al. [306]
Skin lesions	Rainbow trout (<i>Oncorhynchus mykiss</i>)	Markkula et al. [307]
Immune system	Carp (<i>Cyprinus carpio</i>)	Markkula et al. [307]
Cataract	Rainbow trout (<i>Oncorhynchus mykiss</i>)	Markkula et al. [307]
Cataract	Lake trout (<i>Salvelinus namayush</i>)	Roberts [308], Allison [309]
Cataract	Atlantic salmon (<i>Salmo salar</i>)	Wall [310]

The biological weighting function derived for UVB-induced mortality in cod eggs is similar to that reported for DNA [81], which suggests that mortality is a direct result of DNA damage. No evidence was found that UVA had any influence on the UV sensitivity. Calculations based on this biological weighting function indicate that, under current noon surface UV irradiance, 50% of cod eggs located at, or very near below (within 10 cm), the ocean surface would be dead after 42 h of exposure [298]. To counterbalance this high sensitivity to UV exposure (found in the laboratory), many effects actually reduce the solar irradiance on fish eggs: surface water mixing, UV absorption in the water column, cloud cover, etc.

Browman et al. [316] investigated the UV effects on the early life stages of a planktonic Calanoid copepod (*calanus finmarchicus gunnerus*) and of Atlantic cod (*Gadus morhua*). Both are key species for the North Atlantic food webs. The wavelength-specific DNA damage (cyclobutane pyrimidine dimer [CPD] formation per megabase of DNA) is similar to the action spectrum for damage to T7 bacteriophage DNA.

Cases of skin lesions in fish have been reported long ago, as early as 1930. Skin lesions were described as white-to-gray necrotic areas and erosions [297]. As explained by Roberts [308], the sensitivity of fish skin to UV radiation results from the absence of a keratin layer, the poor melanin content, and the presence of dividing cells in all layers of the skin. Therefore, the skin of fish is more sensitive to UV than that of humans [297]. Lesions are typically found on the head, dorso and tail, because these regions are exposed most. Bull-ock [317] wrote a detailed review about skin alterations in fish. Such lesions could be found in Koi carps [306], rainbow trouts [301, 302, 318, 319], brown trout [309], Atlantic salmon [303, 305], and plaice [320]. Exposure of Atlantic salmon to enhanced UVB radiation retarded growth, and decreased the hematocrit value and plasma protein concentration. Furthermore, enhanced UVB radiation affected the plasma immunoglobulin concentration. These results also demonstrated that juvenile Atlantic salmon are not able to fully adapt to increased ambient UVB levels over long-term exposures. The interference with the immune system function suggests a negative effect of UVB on disease resistance [321]. Markkula et al. [322] studied the effect of UVB exposure on common carp (*Cyprinus carpio*),

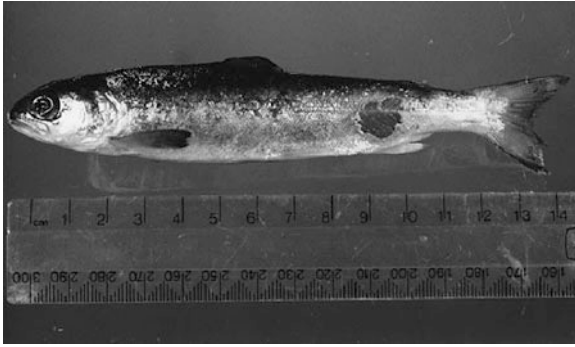
based on the immunomodulation in the blood and head kidney. Their results emphasize the potentially harmful impact of increased solar UVB radiation on the fish immune functions, and show that the effects of short- and long-term exposure differ from each other.

Armstrong et al. [323] investigated larvae of Japanese medaka (*Oryzias latipes*) to determine whether pigmentation modifies the UVB-inducible damage. One-day post-hatch medaka were exposed to UVB radiation, with or without photoreactivating light, 7 h per day for 5 days. At the higher UVB dose tested, wild-type melanophore-containing medaka formed significantly more dimers than at least one of the other strains tested. Wild-type medaka also showed significantly less photorepair capability than the white melanophore-lacking medaka. The presence of melanophores in the wild-type medaka may have contributed to an increased level of tissue damage in this strain when compared to the other strains.

Juvenile rainbow trout were also exposed to UVB using a UV-transparent respirometer chamber. A direct relationship between UVB exposure and the oxygen consumption was observed. Increased swimming activity and restless behavior were also noted [324].

Keeping fish in fish farms with high stocking rates results in a much higher UV exposure [325]. A relationship for UV-induced abnormalities was found for fish farms at high elevation in Bolivia [302] and for different seasons [305, 309].

Sunburn is in many cases the first stage before an outbreak of infectious diseases can be observed [326]. For example, fungal infections commonly occur in fish exposed to excessive sunlight. Outbreaks of *flexibacteriosis maritimus* in cultured Atlantic salmon in Tasmania (Fig. 30) occurred due to high UV exposure during cloud-free days with water temperatures above 21°C [327]. Markkula et al. [307] studied the effect of UVB on the resistance of rainbow trout (*Oncorhynchus mykiss*) against a bacterium (*Yersinia ruckeri*), the causative agent of enteric red mouth disease, and a trematode parasite (*Diplostomum spathaceum*), which causes cataracts in fish. As a result of increased UVB exposure, a significantly higher number of parasites was detected in the eyes of the irradiated fish, indicating reduced resistance against the pathogen. Furthermore, UVB exposure altered the



UV Effects on Living Organisms. Figure 30

Skin lesions of salmonid cutaneous erosion disease caused by *Flexibacter maritimus* in experimentally infected Atlantic salmon (From Handler et al. [327])

immunological and hematological parameters of fish, which also verified the immunomodulatory potential of UVB.

In a retrospective study conducted at 13 salmon farms in Western Ireland, Rodgers [305] showed that economic losses resulting from UV-induced skin damages (sunburn) could be reduced by the use of broadband antibiotics. Six out of 13 farms using antibiotics suffered from such losses, mounting to 10–23% of the post-smolt stocks over the sampling period. In 4 out of the 13 farms, where the cages were shaded, no sunburn cells in the histology of the skin could be observed.

In addition to skin lesions, eye damages can be observed in fish. Cullen et al. [328] exposed rainbow trouts (*Onchorynchus mykiss*) to broadband UV exposure over a period of 205 days. The results support the hypothesis that chronic exposure to ambient levels of UV radiation is cataractogenic. For the Atlantic salmon (*Salmo salar*), Wall [310] could show that the incidence of cataracts was very variable between sites (Ireland, Scotland, and Norway). Whereas in some farms 95% of the salmon had cataract, this fraction was less than 5% in other farms. Blind fish could often be seen near the surface. The pattern of these outbreaks suggested that the diet could be involved, beside other factors like UV exposure. For juvenile scalloped hammerhead sharks (*Sphyrna lewini*), the absorption in their corneal tissues, particularly at wavelengths below 310 nm, increases proportionally to UV exposure [329].

Cataract could also be observed in fish [308]. Allison [309] reported on trout kept in open but covered basins. In these sun-protected basins, 6% of the brown trouts developed cataract, whereas it was as much as 22% in uncovered basins.

To study repair mechanisms and the effects of UV radiation in animals, the Zebra fish is becoming an important model system. Their embryos are easily treated with UV radiation. The DNA damage repair pathways appear to be conserved in Zebra fish and mammals [330].

Due to UV exposure, aquatic organisms also developed protection measures and avoidance strategies. For them, a variety of physiological and biochemical mechanisms are available for reducing the damage by UV exposure. Screening mechanisms include both physical and chemical protection measures. Physical methods are UV absorption of mucus, sporopollenin, and multiple cell walls. Chemical screening methods include UV absorption by mycosporines, mycosporine-like amino acids, scytonemin, 3-Hydroxykynurenin, melanin, and fluorescent pigments. When screening does not eliminate penetration of UVR into the cell, there is a variety of quenching and repair processes available to overcome damage to sensitive cellular components (antioxidants, carotenoids, and the xanthophyll cycle). Damaged proteins are usually replaced by de novo synthesis (turnover), whereas damaged DNA is generally repaired. Several different mechanisms exist for repairing DNA, including photoreactivation, nucleotide excision repair, dimer bypass, and recombinational repair [331].

Coral reef fish were recently discovered to have ultraviolet (UV) radiation screening compounds, most commonly known as mycosporine-like amino acids (MAAs), in their external body mucus. However, little is known about the identity and quantity of MAAs in the mucus of reef fish, or what factors affect their abundance and distribution. In seven species of reef fish (*Labroides dimidiatus* and *Thalassoma lunare* [Labridae]; *Chlorurus sordidus*, *Scarus flavipectoralis*, *S. niger*, *S. rivulatus*, and *S. schlegelii* [Scaridae]) from Queensland, Australia, the UV absorbance of mucus was measured. The mucus of all fish investigated contained MAAs. Depending on species, different combinations and quantities of the MAAs asterina-330, palythene, and mycosporine-N-methylamine

serine were present [332]. Zamzow [333] could show that the UV protective agent in the mucus depends on the level of UV exposure. For shallow-water Caribbean fish (*Scarus iseri* and *Halichoeres bivittatus*), different ways of coping with UV radiation were discovered: *H. bivittatus* shifted the spectral quality of its mucus to increase absorbance of shorter, more damaging wavelengths, whereas *S. iseri* increased the overall UV absorbance of its mucus, with minimal spectral shifting. The influence of the diet and gender on the UV protection by the mucus could be shown for tropical wrasse, *Thalassoma duperrey* [334]. Females had higher rates of skin damage than males. Females sequester UV absorbing compounds in their pelagic

eggs as well as their epithelial mucus, whereas males do not sequester these compounds in the testes.

Behavioral avoidance strategies (Table 12) are not only protection measures, but also a way to reduce UV exposure. Both laboratory and field experiments have demonstrated that many species are negatively phototactic to UV. In addition, UV photoreceptors were found in a variety of fish and invertebrates, suggesting that UV vision may be prominent in aquatic organisms [335–337]. These UV photoreceptors are thought to be used for navigation, communication, enhanced foraging, and possibly UV radiation avoidance. Given the presence of negative phototactic behaviors as well as UV vision, UV radiation may be an important factor

UV Effects on Living Organisms. Table 12 Behavioral avoidance of aquatic organism caused by UV radiation in laboratory (L) and field (F) experiments

Organism	Behavioral activity in the presence of UV radiation	Laboratory/field study
Marine cyanobacteria <i>Microcoleus chthonoplastes</i>	Migrate to greater depths	L
Ciliate <i>Btepharisma japonicum</i>	Backward swimming	L
Cladocerans	Negatively phototactic	L
<i>Daphnia magna</i>	Negatively phototactic	L
Cyclopoid <i>Cyclops serrulatus</i>	Avoid UV exposure	L
Marine echinoid larva <i>Dendraster excentricus</i>	Avoid UV exposure	L
Oplophorid shrimp <i>Systellaspis debilis</i>	Pitching, changing swimming speed, moving the feeding appendages	L
Zebra mussel <i>Dreissena polymorpha</i>	Ceased all swimming and crawling motions	L
<i>D. pulicaria</i>	Rapidly descended from the surface waters	F
<i>D. catawba</i>	Negatively phototactic	F
<i>D. catawba</i>	Negatively phototactic	F
Copepod <i>Diaptomus nevadensis</i>	Negatively phototactic	F
Marine copepod <i>Acartia tonsa</i>	Negatively phototactic	F
Cladoceran <i>Daphnia magna</i>	Negatively phototactic	F
Hydromedusan <i>Polyorchis penicillatus</i>	Negatively phototactic	F
Juvenile rainbow trout (<i>Oncorhynchus mykiss</i>)	Increased swimming activity	F
Juvenile coho salmon (<i>Oncorhynchus kisutch</i>)	Negatively phototactic	L/F

Sources: Leech and Johnsen [338], Alemanni et al. [324], Holtby and Bothwell [340]

influencing migration and abundance patterns, as well as predator-prey and intraspecific interactions [338].

UV vision has been documented in a variety of terrestrial organisms including insects, birds, amphibians, reptiles, and mammals. It is therefore not surprising that many aquatic organisms also perceive light in the UV range. Most UV photoreceptors in aquatic organisms have been described in fish species; however, UV photoreceptors have also been reported in bacteria and algae as well as in some species of protozoan, annelids, cnidarians, and crustaceans. Many UV photoreceptors have a maximum absorbance peak in the UVA range, although UVB photoreceptors have also been documented in some species. One explanation for the rarity of UVB vision is that UVB radiation is potentially more damaging to the eye [338, 339].

Vitamin D Fish store large quantities of vitamin D in their liver and fat tissues, including the fat associated with muscles, and this makes fish an important dietary source of vitamin D for their predators or for humans. Plankton, the chief food source of fish, was assessed [341] as the possible dietary origin of vitamin D in fish. High amounts of provitamins D and vitamins D (D_2 and D_3) were found in freshwater plankton [342] during summer months. Thus, plankton may be an important contributor to vitamin D in fish.

For some species of fish (rainbow trout and Mozambique tilapia) photosynthesis of vitamin D_3 from 7-DHC was reported [343]. However, it is unlikely that the photosynthesis of vitamin D_3 plays a significant role under natural conditions, where fish are less exposed to UV light.

The role of vitamin D in fish physiology is still enigmatic [344]. Until the 1970s, the consensus was that fish accumulated but not metabolized vitamin D. Now, 4 decades later, there is substantial evidence that fish have a vitamin D endocrine system with similar functions as in mammals.

Zoo Animals

The geographical origins of many exotic animals are areas with high-UV intensity. When brought into zoological gardens, they might suffer under a deficiency of UV exposure, especially at mid- and high-latitudes. On the other hand, increased UV

exposure may cause photodamage of the skin and the eyes. To provide these animals with an appropriate environment, their solar UV exposure has to be taken into account.

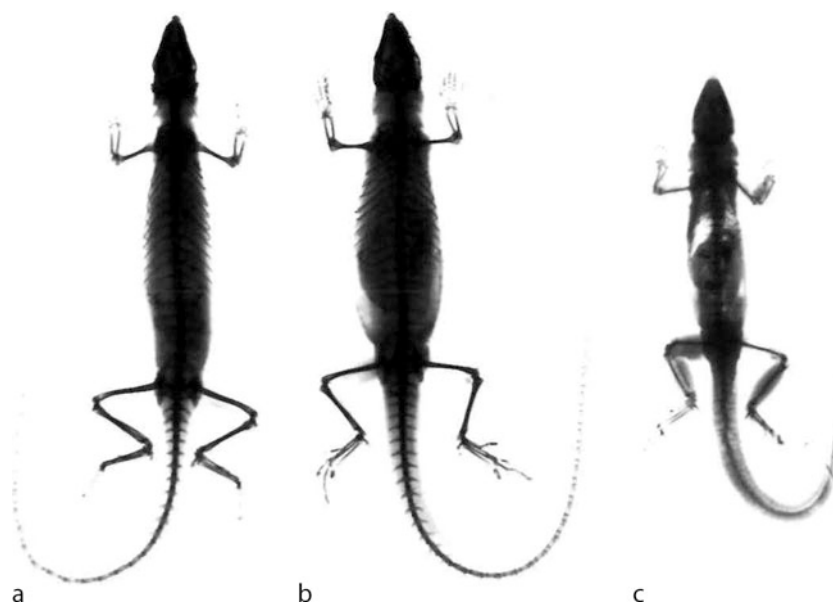
Lack of UV radiation strongly reduces the availability of vitamin D_3 . Therefore, animals can suffer from rickets (rickets) and osteodystrophy, which causes irreversible bone degeneration of the locomotion system and substantial deformation of the jaw. This handicaps both their movement and food intake. Vitamin D can also be supplied by feed. However, excessive dietary vitamin D can result in toxic effects and death. Conversely, high-UV exposure does not result in excessive vitamin D levels [345].

For some species, the vitamin D supply can be substituted by vitamin D_2 supplementation in the feed. However, Göltenboth [346] illustrated that rickets in primates could not be healed by such vitamin D_2 supplementation. The symptoms could only be eradicated by exposure to lamps with a high UV output (Ultravitalux-lamps, Osram). Several species of New World primates have an absolute requirement for vitamin D_3 , either from the diet or through exposure to UV radiation. They seem to metabolize vitamin D_2 too quickly to maintain adequate serum levels of 1,25-dihydroxyergocalciferol [347]. Hofmann-Parisot [348] described typical differences in diverse primate species concerning the vitamin D metabolism.

Tonge [349] has shown that the application of supplemental UV radiation can prevent the deformation of the tail of lizards. Bosch and Frank [350] pointed out that in terrariums, 11% of turtles and 7% of iguanas display a vitamin D deficiency syndrome.

McCrystal and Behler [351] and Townsend and Cole [352] indicated that for the breeding success in lizards, a sufficient supply in UV is necessary. This can be attributed to interactions with the calcium metabolism, because calcium is responsible for the mechanical stability of their eggs (Fig. 31).

Another fundamental problem for lizards raised in captivity may be that their viable eggs incubate full term but then fail to hatch, even when produced by apparently healthy adults. The fully developed, dead embryos appear normal but have poorly mineralized skeletons. This problem is similar to that of chicken, and can also be corrected by supplementing UV to the adult female prior to oviposition [284, 285].



UV Effects on Living Organisms. Figure 31

X-ray image of whiptail lizards (*Cnemidophorus exsanguis*). (a) Healthy, wild-caught adult. (b) Healthy laboratory-reared adult. (c) Laboratory-reared young adult with metabolic bone disease (Source: Townsend and Cole [352])

Turtles receive vitamin D either from dietary or from UV exposure. It was shown by Acierno et al. [353] that UV is much more effective than diet. In 1997, Manning and Grigg [47] found that basking in a freshwater turtle is longer than thermoregulation would require, which indicates that turtles regulate their vitamin D status by basking.

For snakes and turtles, Acierno et al. [354] investigated the vitamin D supply following 28 days of additional UV exposure. As a result of such an additional UV exposure, the plasma concentration of 25-hydroxyvitamin D₃ was about 3.4 times higher than the control group in snakes, and 2.3 times higher in turtles.

For various lizards, snakes, and chameleons [50, 345], it could be shown that the exposure behavior to natural sunlight depends on the vitamin D status. Animals with low dietary vitamin D₃ intake significantly increased their spontaneous exposure to solar radiation, compared to those with a high dietary vitamin D₃ intake. These results demonstrate the importance of basking for nonthermoregulatory purposes and, more specifically, that basking constitutes an integral mechanism for the regulation of a vital hormone,

vitamin D₃. Ferguson et al. [49] found some evidence that the vitamin D photosynthesis in the skin may reflect a correlated response associated with the need of vitamin D. However, the possibility that specific adaptations for the photosynthesis of vitamin D and for the protection from skin damage could involve independent mechanisms still needs more investigation.

Ferguson et al. [48, 49] have shown that Panther chameleons (*Furcifer pardalis*) can regulate UV exposure behaviorally or by a change in skin color, so that their vitamin D production is enhanced at low UV levels. As Karsten et al. [50] have shown, Panther chameleons and possibly reptiles, in general, bask also for reasons other than thermoregulation. In dependence of their dietary vitamin D₃ intake, Panther chameleons behaviorally regulate their exposure to solar UV to ensure an optimal vitamin D₃ status with high precision. Their UVA-sensitive retina is helpful in this respect. Vitamin D supplementation in lizards is not easy. It may become toxic, and possibly lethal [284].

Hofmann-Parisot [348] investigated the use of artificial UV sources that were used in keeping zoo animals healthy. It was found that ≈60% of the zoo gardens surveyed apply artificial UV radiation to animals.

Of that number, 48% use UV only for treating reptiles, and 52% do this also for mammals and avian. Interestingly, this survey also provided a very fuzzy picture of the UV supplementation practice: UV intensity and exposition times, types of lamps, and radiation geometry were highly variable, illustrating that no sufficient information is available concerning the UV needs of specific animal species. As illustrated by the case of chameleons, animal behavior has to be taken into account.

Gehrmann [355] realized the importance of using UV sources in animal husbandries and investigated different types of lamps, as well as the reflection and transmission properties of some materials that are used in zoos [356]. (Note that, in this kind of study, the spectral distribution of these UV sources and the specific animal requirements would have to be considered as well.)

Even though it is well known that most of the observed diseases related to vitamin D deficits are caused by a lack of UV exposure, no systematic investigations have been performed to optimize the UV irradiance or the exposure time for various animals. Most of the applications of artificial UV sources resulted from “educated guess.” For snakes and lizards, however, Ferguson et al. [345] estimated the natural UV irradiance by using the UV index of their natural environment. The conversion rate of vitamin D due to UV exposure was investigated by Ferguson et al. [285] for panther chameleons (*Furcifer pardalis*).

Although insufficient natural UV radiation is typical in zoo animals, overexposure to artificial UV can also appear, if inappropriate UV sources are in use. For a ball python (*Python regius*) and a blue tongue skink (*Tiliqua nigrolutea*), the application of an otherwise appropriate UV lamp, but with an extremely high UVB output, showed epidermal erosion and ulceration, with severe epidermal basal cell degeneration and necrosis, and superficial dermatitis as well as bilateral ulcerative keratoconjunctivitis with bacterial colonization [357]. Veterinarians who are consulted for reptiles with ocular and/or cutaneous disease of unapparent cause should fully evaluate the specifics of the vivarium light sources. In general, overexposure to UV radiation can cause eye and skin damage, skin cancer, and poor reproduction in reptiles and amphibians [345].

Future Directions

The so-called ozone hole was detected in ozone measurements of the late 1970s [358]. The reduction of the total ozone content in the stratosphere and the resulting decrease of absorbing capacity for UV radiation [359] cause a hype in research on the biological effects of UV radiation. Up to this moment the solar radiation and therefore also the spectral range of UV radiation was assumed to be constant without remarkable variability. In the following 2 decades, a steep increasing number of investigations were published which were concentrated primarily to the monocausal effects of UV radiation on plants, animals, and humans.

These research activities of biologically effective UV radiation due to ozone depletion were replaced by the climate change discussion starting in the late 1990s. By more sophisticated general circulation models (GCM) also parameters like solar radiation were predicted for the next century. This leads to the opportunity to predict the biologically effective UV radiation as well. These climate change scenarios for a century were the basis for a more holistic approach to describe the interaction between UV radiation and flora as well as fauna. The coupling of the biosphere with GCM can be used to assess effects of the climate change on the biosphere and vice versa (e.g., reduced cloud cover leads to a higher UV exposure of plankton which reduces the plankton reproduction, which causes a reduction of carbon storage and finally an increase of CO₂ release, influencing the radiative forcing as an important input parameter of GCM). For humans the resulting migration scenarios and the adaptation of the behavior to the changing climate can be used to calculate the modification of the UV skin exposure and the health implications.

Beside UV effects, which cause a damage (e.g., eye, skin), also beneficial effects are under discussion, which are correlated with the vitamin D synthesis due to UV radiation. A sufficient supply with vitamin D is necessary for skeletal health. Further it is assumed as protective measure against several kinds of cancer in humans. For these investigations an appropriate animal model would be a great help.

For several UV effects different working hypothesis are in use, which are verified very rarely. First of all the assumption of reciprocity is a simple prerequisite

which says that the effect is depending on the UV dose, which is calculated by the intensity of the radiation times the exposure time. Secondly we assume that a polychromatic UV exposure is caused by an unweighted sum of all spectral UV doses. This means that the weighting of the UV radiation and the resulting biologically effective UV radiation can be described as photoadditive, without photoprotection or photoaugmentation due to certain wavelength ranges [360].

Bibliography

Primary Literature

- Bunsen RW, Roscoe HE (1857) Photochemische Untersuchungen. Poggendorfs Ann 100:51
- Schindl A, Rosado-Schlösser B, Trautinger F (2001) Die Reziprozitätsregel in der Photobiologie. Eine Übersicht Der Hautarzt 52:779–785
- Diffey BL, Kochevar IE (2007) Basic principles of photobiology. In: Lim HW, Hönigsmann H, Hawk JL (eds) Photodermatology. Informa Healthcare USA, New York, pp 15–27
- Schwarzschild K (1900) On the law of reciprocity for bromide of silver gelatin. Astrophys J 11:89
- Martin JW, Chin JW, Nguyen T (2003) Reciprocity law experiments in polymeric photodegradation: a critical review. Prog Org Coat 47:292–311
- Reischl S, Henrich J, Schauburger G (2000) Time course and dose-response-relationship of erythema of the pigs induced by UVB- and UVC-radiation. In: 13th international congress on photobiology, San Francisco, 1–6 July 2000, p 136
- Farr PM, Besag JE, Diffey BL (1988) The time course of UVB and UVC erythema. J Invest Dermatol 91:454–457
- Mehlhorn G, Steiger A (1974) Künstliche UV-Strahlung in der Tierproduktion. Gustav Fischer Verlag, Jena
- Kasper I (1986) Abschätzung biologischer UV-Wirkungen durch Bestrahlung von Großtieren mit künstlichen UV-Quellen. Dissertation Veterinärmedizinische Universität Wien, Wien
- MacKinley AF, Diffey BL (1987) A reference action spectrum for ultraviolet induced erythema in human skin. CIE J 6:17–22
- Gates FL (1930) A study of the bactericidal action of ultra violet light. III. The absorption of ultra violet light by bacteria. J Gen Physiol 14:31–42
- Hollaender A, Emmons CW (1941) Wavelength dependence of mutation production in the ultraviolet with special emphasis on fungi. Cold Spring Harb Symp Quant Biol 9:179–185
- Herschel W (1800) Investigation of the powers of the prismatic colours to heat and illuminate objects; with remarks, that prove the different refrangibility of radiant heat. To which is added, an inquiry into the method of viewing the sun advantageously, with telescopes of large apertures and high magnifying powers. Philos Trans R Soc Lond 90:255–283
- Herschel W (1800) Experiments on the refrangibility of the invisible rays of the sun. Philos Trans R Soc Lond 90:284–292
- Ritter JW (1801) Ann Phys 7:527
- Daubeny C (1836) On the action of light upon plants and plants upon the atmosphere. Philos Trans R Soc Lond 126:149–176
- Draper JW (1884) Note on decomposition of carbonic acid by the leaves of plants under the influence of yellow light. Philos Mag 25:159–173
- Engelmann TW (1882) Über Sauerstoffausscheidung von Pflanzenzellen im Microspectrum. Bot Zeit 40:419–426
- Hausser KW, Vahle W (1921) Die Abhängigkeit des Lichterythems und der Pigmentbildung von der Schwingungszahl (Wellenlänge) der erregenden Strahlung. Strahlentherapie 13:41–71
- Coblentz WW, Stair R (1934) Data on the spectral erythemic reaction of the untanned human skin to ultraviolet radiation. Bur Stand J Res 12:13–14
- Luckiesh M, Holladay LL, Taylor AH (1930) Reaction of untanned human skin to ultraviolet radiation. J Opt Soc Am 20:423–426
- Hausser KW, Vahle W (1927) Sonnenbrand und Sonnenbräunung. Wiss Veröffl Siemens Konzern 6:101–120
- CIE (1937) Neuvieme Session Berlin und Karlsruhe Juillet 1935. Cambridge University Press, Cambridge, UK
- McKinley AF, Diffey BL (1987) A reference action spectrum for ultraviolet induced erythema in human skin. CIE J 6:17–22
- Duggar BM, Hollaender A (1934) Irradiation of plant viruses and of microorganisms with monochromatic light: I. The virus of typical tobacco mosaic and serratia marcescens as influenced by ultraviolet and visible light. J Bacteriol 27: 219–239
- Gates FL (1934) Results of irradiating staphylococcus aureus bacteriophage with monochromatic ultraviolet light. J Exp Med 60:179–188
- Sonne C, Rekling E (1927) Behandlung experimenteller Rattenrachitis mit monochromatischem ultravioletten Licht. Strahlentherapie 35:552–558
- Henschke U, Schulze R (1939) Über Pigmentierung durch langwelliges Ultraviolett. Strahlentherapie 65:14–42
- Reiter T, Gabor D (1928) Ultraviolette Strahlen und Zellteilung. Strahlentherapie 28:125–132
- Blumthaler M, Ambach W (1988) Solar UV-B albedo of various surfaces. Photochem Photobiol 54:429–432
- Monteith JL (1973) Principles of environmental physics. Arnold, London
- Quaschnig V, Hanitsch R (1998) Irradiance calculation on shaded surfaces. Sol Energy 62:369–375
- Blumthaler M, Ambach W, Ellinger R (1996) UV-Bestrahlung von horizontalen und vertikalen Flächen im Hochgebirge. Wetter Leben 1:25–31
- Schauburger G (1990) Model for the global irradiance of the solar biologically-effective ultraviolet radiation on inclined surfaces. Photochem Photobiol 52:1029–1032

35. Grant RH (1998) Ultraviolet irradiance of inclined planes at the top of the plant canopies. *Agric For Meteorol* 89:281–293
36. Webb AR, Weihs P, Blumthaler M (1999) Spectral UV irradiance on vertical surfaces: a case study. *Photochem Photobiol* 69:464–470
37. Hay JE (1979) Calculation of monthly mean solar radiation for horizontal and inclined surfaces. *Sol Energy* 23:301–307
38. Schaubberger G (1992) Anisotropic model for the diffuse biologically effective irradiance of solar UV radiation on inclined surfaces. *Theor Appl Climatol* 46:45–51
39. Klucher TM (1979) Evaluation of models to predict insolation on tilted surfaces. *Sol Energy* 26:179–184
40. Hay JE, McKay DC (1985) Estimating solar irradiance on inclined surfaces: a review and assessment of methodologies. *Int J Sol Energy* 3:203–240
41. Reindl DT, Beckmann WA, Duffie JA (1990) Diffuse fraction correlations. *Sol Energy* 45:1–7
42. Utrillas MP, Marin MJ, Esteve AR, Estelles V, Tena F, Canada J et al (2009) Diffuse ultraviolet erythral irradiance on inclined planes: a comparison of experimental and modeled data. *Photochem Photobiol* 85:1245–1253
43. Mech M, Koepke P (2004) Model for UV irradiance on arbitrarily oriented surfaces. *Theor Appl Climatol* 77:151–158
44. Hess M, Koepke P (2008) Modelling UV irradiances on arbitrarily oriented surfaces: effects of sky obstructions. *Atmos Chem Phys* 8:3583–3591
45. Dirmhirn I (1964) *Das Strahlungsfeld im Lebensraum*. Akademische Verlagsgesellschaft, Frankfurt
46. Geiger R (1961) *Das Klima der bodennahen Luftschicht*, 4. Aufl. ed. Vieweg, Braunschweig
47. Manning B, Grigg GC (1997) Basking is not of thermoregulatory significance in the “basking” freshwater turtle *Emyda signata*. *Copeia* 1997:579–584
48. Ferguson GW, Gehrmann WH, Karsten KB, Hammack SH, McRae M, Chen TC et al (2003) Do panther chameleons bask to regulate endogenous vitamin D₃ production? *Physiol Biochem Zool* 76:52–59
49. Ferguson GW, Gehrmann WH, Karsten KB, Landwer AJ, Carman EN, Chen TC et al (2005) Ultraviolet exposure and vitamin D synthesis in a sun-dwelling and a shade-dwelling species of *Anolis*: are there adaptations for lower ultraviolet B and dietary vitamin D₃ availability in the shade? *Physiol Biochem Zool* 78:193–200
50. Karsten KB, Ferguson GW, Chen T, Holick MF (2009) Panther chameleons, *furcifer pardalis*, behaviorally regulate optimal exposure to UV depending on dietary vitamin D status. *Physiol Biochem Zool* 82:218–225
51. Bidmon HJ, Stumpf WE (1996) Vitamin D target systems in the brain of the green lizard *Anolis carolinensis*. *Anat Embryol* 193:145–160
52. Bowmaker JK, Loew ER, Ott M (2005) The cone photoreceptors and visual pigments of chameleons. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 191:925–932
53. Tovee MJ (1995) Ultra-violet photoreceptors in the animal kingdom: their distribution and function. *Trends Ecol Evol* 10:455–460
54. Bianca W, Wegmann H (1974) Die Durchlässigkeit von Rinderhaaren für ultraviolette Strahlung. *Schweiz Arch Tierheilk* 116:114–146
55. Keck G, Kasper I, Schaubberger G, Cabaj A (1987) Biological effects of UV-irradiation on horses with artificial UV-sources. In: Passchier WF, Bosnjakovic BFM (eds) *Human exposure to ultraviolet radiation – risk and regulations*. Elsevier, Amsterdam
56. Keck G, Kasper I, Schaubberger G, Cabaj A (1988) UV-Wirkungen in Pferdesolarien? *Berl Münchner Tierärztl Wochenschr* 101:49–52
57. Chapman RS, Cooper KD, De Fabo EC, Frederick JE, Gelatt KN, Hammond SP et al (1995) Solar ultraviolet radiation and the risk of infectious disease: summary of a workshop. *Photochem Photobiol* 61:223–247
58. Schaubberger G (1997) Biological relevance of solar radiation to animals. *Wien Tierärztl Mschr* 84:2–13
59. Wolfe WL, Zissig GJ (1978) *The infrared handbook*. Office of Naval Research, Department of the Navy, Washington
60. Chadysiene R, Girgzdys A (2008) Ultraviolet radiation albedo of natural surfaces. *J Environ Eng Landscape Manage* 16:83–88
61. Morris DP, Zagarese H, Williamson CE, Balseiro EG, Hargreaves BR, Modenutti B et al (1995) The attenuation of solar UV radiation in lakes and the role of dissolved organic carbon. *Limnol Oceanogr* 40:1381–1391
62. Reitner B, Herndl GJ, Herzig A (1997) Role of ultraviolet-B radiation on photochemical and microbial oxygen consumption in a humic-rich shallow lake. *Limnol Oceanogr* 42:950–960
63. Booth CR, Morrow JH (1997) The penetration of UV into natural waters. *Photochem Photobiol* 65:254–257
64. Morris DP, Hargreaves BR (1997) The role of photochemical degradation of dissolved organic carbon in regulating the UV transparency of three lakes on the Pocono Plateau. *Limnol Oceanogr* 42:239–249
65. Graneli W, Lindell M, Tranvik L (1996) Photo-oxidative production of dissolved inorganic carbon in lakes of different humic content. *Limnol Oceanogr* 41:698–706
66. Markager S, Vincent WF (2000) Spectral light attenuation and the absorption of UV and blue light in natural waters. *Limnol Oceanogr* 45:642–650
67. Vincent WF, Rae R, Laurion I, Howard-Williams C, Priscu JC (1998) Transparency of Antarctic ice-covered lakes to solar UV radiation. *Limnol Oceanogr* 43:618–624
68. Sommaruga R, Psenner R (1997) Ultraviolet radiation in a high mountain lake of the Austrian Alps: air and underwater measurements. *Photochem Photobiol* 65:957–963
69. Sommaruga R, Augustin G (2006) Seasonality in UV transparency of an alpine lake is associated to changes in phytoplankton biomass. *Aquat Sci* 68:129–141
70. Huovinen PS, Penttilä H, Soimasuo MR (2003) Spectral attenuation of solar ultraviolet radiation in humic lakes in Central Finland. *Chemosphere* 51:205–214

71. Smith RC, Baker KS (1981) Optical properties of the clearest natural waters (200–800 nm). *Appl Opt* 20:177–184
72. Baker KS, Smith RC (1982) Bio-optical classification and model of natural waters. *Limnol Oceanogr* 27:500–509
73. De Lange HJ (2000) The attenuation of ultraviolet and visible radiation in Dutch inland waters. *Aquat Ecol* 34:215–226
74. Bukaveckas PA, Robbins-Forbes M (2000) Role of dissolved organic carbon in the attenuation of photosynthetically active and ultraviolet radiation in Adirondack lakes. *Freshw Biol* 43:339–354
75. Squires MM, Lesack LFW (2003) Spatial and temporal patterns of light attenuation among lakes of the Mackenzie Delta. *Freshw Biol* 48:1–20
76. Sommaruga R (2001) The role of solar UV radiation in the ecology of alpine lakes. *J Photochem Photobiol B* 62:35–42
77. Smith RC, Baker KS (1978) Optical classification of natural waters. *Limnol Oceanogr* 23:260–267
78. Smith RC, Baker KS (1979) Penetration of UV-B and biologically effective close-rates in natural waters. *Photochem Photobiol* 29:311–323
79. Jerlov NG (1968) *Optical oceanography*. Elsevier, Amsterdam
80. Williamson CE, Stemberger RS, Morris DP, Frost TM, Paulsen SG (1996) Ultraviolet radiation in North American lakes: attenuation estimates from DOC measurements and implications for plankton communities. *Limnol Oceanogr* 41:1024–1034
81. Setlow RB (1974) The wavelengths in sunlight effective in producing skin cancer: a theoretical analysis. *Proc Natl Acad Sci USA* 71:3363–3366
82. Grenfell TC, Warren SG, Mullen PC (1994) Reflection of solar radiation by the Antarctic snow surface at ultraviolet, visible, and near-infrared wavelengths. *J Geophys Res* 99: 18669–18684
83. Blumthaler M, Ambach W (1988) Solar UVB-albedo of various surfaces. *Photochem Photobiol* 48:85–88
84. Meinander O, Kontu A, Lakkala K, Heikkilä A, Ylianttila L, Toikka M (2008) Diurnal variations in the UV albedo of arctic snow. *Atmos Chem Phys* 8:6551–6563
85. Perovich DK (1993) A theoretical model of ultraviolet light transmission through Antarctic sea ice. *J Geophys Res* 98:22579–22588
86. Cockell CS, Rettberg P, Horneck G, Wynn-Williams DD, Scherer K, Gugg-Helminger A (2002) Influence of ice and snow covers on the UV exposure of terrestrial microbial communities: dosimetric studies. *J Photochem Photobiol B* 68:23–32
87. Warren SG, Brandt RE, Grenfell TC (2006) Visible and near-ultraviolet absorption spectrum of ice from transmission of solar radiation into snow. *Appl Opt* 45:5320–5334
88. King MD, France JL, Fisher FN, Beine HJ (2005) Measurement and modelling of UV radiation penetration and photolysis rates of nitrate and hydrogen peroxide in Antarctic sea ice: an estimate of the production rate of hydroxyl radicals in first-year sea ice. *J Photochem Photobiol A* 176:39–49
89. Perovich DK (2001) UV radiation and the optical properties of sea ice and snow. In: Hesse D (ed) *UV radiation and Arctic ecosystems*, pp 73–89
90. Lesser MP, Lamare MD, Barker MF (2004) Transmission of ultraviolet radiation through the Antarctic annual sea ice and its biological effects on sea urchin embryos. *Limnol Oceanogr* 49:1957–1963
91. Fritsen CH, Priscu JC (1999) Seasonal change in the optical properties of the permanent ice cover on Lake Bonney, Antarctica: consequences for lake productivity and phytoplankton dynamics. *Limnol Oceanogr* 44:447–454
92. Grenfell TC, Maykutt GA (1977) The optical properties of ice and snow in the Arctic basin. *J Glaciol* 18:445–463
93. Belzile C, Gibson JAE, Vincent WF (2002) Colored dissolved organic matter and dissolved organic carbon exclusion from lake ice: implications for irradiance transmission and carbon cycling. *Limnol Oceanogr* 47:1283–1293
94. Belzile C, Johannessen SC, Gosselin M, Demers S, Miller WL (2000) Ultraviolet attenuation by dissolved and particulate constituents of first-year ice during late spring in an Arctic polynya. *Limnol Oceanogr* 45:1265–1273
95. Kepner JRL, Wharton RA Jr, Collier RD, Cockell CS, Jeffrey WH (2000) UV radiation and potential biological effects beneath the perennial ice cover of an antarctic lake. *Hydrobiologia* 427:155–165
96. Felip M, Camarero L, Catalan J (1999) Temporal changes of microbial assemblages in the ice and snow cover of a high mountain lake. *Limnol Oceanogr* 44:973–987
97. Felip M, Wille A, Sattler B, Psenner R (2002) Microbial communities in the winter cover and the water column of an alpine lake: system connectivity and uncoupling. *Aquat Microb Ecol* 29:123–134
98. Savage DC (1977) Microbial ecology of the gastrointestinal tract. *Annu Rev Microbiol* 31:107–133
99. Berg RD (1996) The indigenous gastrointestinal microflora. *Trends Microbiol* 4:430–435
100. Cristiani H (1893) Analyse bacteriologique de l'air des hauteurs, puis un voyage en ballon. *l'Inst Pasteur* 7:665–671
101. Harz CO (1904) Bakteriologische Untersuchungen der freien Atmosphäre mittels Luftballons nebst Bemerkungen über den atmosphärischen Staub. *Fb dtsh Luftsch Verb* 1904:293–302
102. Rogers LA, Meier FC (1936) U.S. Army Air Corps stratosphere flight of 1935 in the balloon "Explorer II." Technical papers, The National Geographic Society, Washington, DC, pp 146–151
103. Imshenetsky AA, Lysenko SV, Kazakov GA (1978) Upper boundary of the biosphere. *Appl Environ Microbiol* 35:1–5
104. Fulton JD (1966) Microorganisms of the upper atmosphere. 3. Relationship between altitude and micropopulation. *Appl Microbiol* 14:237–240
105. Yang Y, Itahashi S, Yokobori S, Yamagishi A (2008) UV-resistant bacteria isolated from upper troposphere and lower stratosphere. *Biol Sci Space* 22:18–25
106. Tong Y, Che F, Xu X, Chen M, Ye B (1993) Population study of atmospheric bacteria at the Fengtai district of Beijing on two representative days. *Aerobiologia* 9:69–74
107. Meier FC, Lindbergh CA (1935) Collecting micro-organisms from the Arctic atmosphere. *Sci Mon* 40:5–20

108. Griffin DW, Garrison VH, Herman JR, Shinn EA (2001) African desert dust in the Caribbean atmosphere: microbiology and public health. *Aerobiologia* 17:203–213
109. Griffin DW, Kubilay N, Koçak M, Gray MA, Borden TC, Shinn EA (2007) Airborne desert dust and aeromicrobiology over the Turkish Mediterranean coastline. *Atmos Environ* 41:4050–4062
110. Griffin DW, Westphal DL, Gray MA (2006) Airborne microorganisms in the African desert dust corridor over the mid-Atlantic ridge, Ocean Drilling Program, Leg 209. *Aerobiologia* 22:211–226
111. Mims SA, Mims FM III (2004) Fungal spores are transported long distances in smoke from biomass fires. *Atmos Environ* 38:651–655
112. Brown JKM, Hovmoll MS (2002) Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science* 297:537–541
113. Tong Y, Lighthart B (1997) Solar radiation is shown to select for pigmented bacteria in the ambient outdoor atmosphere. *Photochem Photobiol* 65:103–106
114. Reiling F (2000) Dissemination of bacteria from the mouth during speaking, coughing, and otherwise. *JAMA* 284:156
115. Giese AC (1976) *Living with our sun's ultraviolet rays*. Plenum, New York
116. Mims FM III (2005) Avian influenza and UV-B blocked by biomass smoke. *Environ Health Perspect* 113(12):A806–A807
117. Coffin RB, Velinsky DJ, Devereux R, Price WA, Cifuentes LA (1990) Stable carbon isotope analysis of nucleic acids to trace sources of dissolved substrates used by estuarine bacteria. *Appl Environ Microbiol* 56:2012–2020
118. Paul JH, Jeffrey WH, DeFlaun M (1985) Particulate DNA in subtropical oceanic and estuarine planktonic environments. *Mar Biol* 90:95–101
119. Paul JH, Carlson DJ (1984) Genetic material in the marine environment: implication for bacterial DNA. *Limnol Oceanogr* 29:1091–1097
120. Fuhrman JA, Sleeter TD, Carlson CA, Proctor LM (1989) Dominance of bacterial biomass in the Sargasso Sea and its ecological implications. *Mar Ecol Prog Ser* 57:207–217
121. Boehme J, Frischer ME, Jiang SC, Kellogg CA, Pichard S, Rose JB et al (1993) Viruses, bacterioplankton, and phytoplankton in the southeastern Gulf of Mexico: distribution and contribution to oceanic DNA pools. *Mar Ecol Prog Ser* 97:1–10
122. Cho BC, Azam F (1990) Biogeochemical significance of bacterial biomass in the ocean's euphotic zone. *Mar Ecol Prog Ser* 63:253–259
123. Azam F, Fenchel T, Field JG, Gray JS, Meyer-Reil LA, ThIngstad F (1983) The ecological role of water-column microbes in the sea. *Mar Ecol Prog Ser* 10:257–263
124. Garcia-Pichel F (1994) A model for internal self-shading in planktonic organisms and its implications for the usefulness of ultraviolet sunscreens. *Limnol Oceanogr* 39:1704–1717
125. Aas P, Lyons MM, Pledger R, Mitchell DL, Jeffrey WH (1996) Inhibition of bacterial activities by solar radiation in nearshore waters and the Gulf of Mexico. *Aquat Microb Ecol* 11:229–238
126. Sieracki ME, Sieburth JM (1986) Sunlight-induced growth delay of planktonic marine bacteria in filtered seawater. *Mar Ecol Prog Ser* 33:19–27
127. Bailey CA, Neihof RA, Tabor PS (1983) Inhibitory effect of solar radiation on amino acid uptake in Chesapeake Bay bacteria. *Appl Environ Microbiol* 46:44–49
128. Herndl GJ, Muller-Niklas G, Frick J (1993) Major role of ultraviolet-B in controlling bacterioplankton growth in the surface layer of the ocean. *Nature* 361:717–719
129. Alonso-Sáez L, Gasol JM (2007) Seasonal variations in the contributions of different bacterial groups to the uptake of low-molecular-weight compounds in Northwestern Mediterranean coastal waters. *Appl Environ Microbiol* 73:3528–3535
130. Häder DP, Kumar HD, Smith RC, Worrest RC (2007) Effects of solar UV radiation on aquatic ecosystems and interactions with climate change. *Photochem Photobiol Sci* 6:267–285
131. Joux F, Jeffrey WH, Lebaron P, Mitchell DL (1999) Marine bacterial isolates display diverse responses to UV-B radiation. *Appl Environ Microbiol* 65:3820–3827
132. Sommaruga R, Obernosterer I, Herndl GJ, Psenner R (1997) Inhibitory effect of solar radiation on thymidine and leucine incorporation by freshwater and marine bacterioplankton. *Appl Environ Microbiol* 63:4178–4184
133. Halac S, Felipe M, Camarero L, Sommaruga-Wögrath S, Psenner R, Catalan J et al (1997) An in situ enclosure experiment to test the solar UVB impact on plankton in a high-altitude mountain lake. I. Lack of effect on phytoplankton species composition and growth. *J Plankton Res* 19:1671–1686
134. Winter C, Moeseneder MM, Herndl GJ (2001) Impact of UV radiation on bacterioplankton community composition. *Appl Environ Microbiol* 67:665–672
135. Jiang H, Dong H, Yu B, Liu X, Li Y, Ji S et al (2007) Microbial response to salinity change in Lake Chaka, a hypersaline lake on Tibetan plateau. *Environ Microbiol* 9:2603–2621
136. Liu Y, Yao T, Jiao N, Kang S, Zeng Y, Huang S (2006) Microbial community structure in moraine lakes and glacial meltwaters, Mount Everest. *FEMS Microbiol Lett* 265:98–105
137. Zhang G, Niu F, Ma X, Liu W, Dong M, Feng H et al (2007) Phylogenetic diversity of bacteria isolates from the Qinghai-Tibet Plateau permafrost region. *Can J Microbiol* 53:1000–1010
138. Zhang G, Ma X, Niu F, Dong M, Feng H, An L et al (2007) Diversity and distribution of alkaliphilic psychrotolerant bacteria in the Qinghai-Tibet Plateau permafrost region. *Extremophiles* 11:415–424
139. Dib J, Motok J, Zenoff VF, Ordóñez O, Farias ME (2008) Occurrence of resistance to antibiotics, UV-B, and arsenic in bacteria isolated from extreme environments in high-altitude (above 4400 m) Andean wetlands. *Curr Microbiol* 56:510–517
140. Fernandez-Zenoff V, Sineriz F, Farias ME (2006) Diverse responses to UV-B radiation and repair mechanisms of bacteria isolated from high-altitude aquatic environments. *Appl Environ Microbiol* 72:7857–7863

141. Ordonez OF, Flores MR, Dib JR, Paz A, Farias ME (2009) Extremophile culture collection from Andean lakes: extreme pristine environments that host a wide diversity of microorganisms with tolerance to UV radiation. *Microb Ecol* 58:461–473
142. Kepner RL Jr, Wharton RA Jr, Suttle CA (1998) Viruses in Antarctic lakes. *Limnol Oceanogr* 43:1754–1761
143. Downes A, Blunt TP (1877) Researches on the effect of light upon bacteria and other organisms. *Proc R Soc Med* 26: 488–500
144. Hertel E (1904) Ueber Beeinflussung des Organismus durch Licht, speziell durch die chemisch wirksamen Strahlen. *Z Allg Physiologie* 4:1–43
145. Hertel E (1905) Ueber physiologische Wirkung von Strahlen verschiedener Wellenlänge. *Z Allg Physiologie* 5:95–122
146. Hockberger PE (2002) A history of ultraviolet photobiology for humans, animals and microorganisms. *Photochem Photobiol* 76:561–579
147. Cabaj A, Sommer R, Pribil W, Haider T (2002) The spectral UV sensitivity of microorganisms used in biosimetry. *Water Sci Technol Water Supply* 2(3):175–181
148. Rauth AM (1965) The physical state of viral nucleic acid and the sensitivity of viruses to ultraviolet light. *Biophys J* 5:257–273
149. EPA (1999) Alternative disinfectants and oxidants guidance manual. EPA 815-R-99-014, United States Environmental Protection Agency, Cincinnati
150. Solsona F, Méndez JP (2003) Water disinfection. PAHO/CEPIS/PUB/03.89
151. WHO (2006) Guidelines for safe recreational water environments, vol 2, Swimming pools and similar environments. WHO, Geneva
152. WHO (2003) Guidelines for safe recreational water environments, vol 1, Coastal and fresh waters. WHO, Geneva
153. Henry V, Helbronner A, Recklinghausen M (1910) Nouvelles recherches sur la sterilization de grandes quantités d'eau par les rayons ultraviolets. *Comp Rend Acad Sci* 151:677–680
154. Wegelin M, Canonica S, Mechsner K, Fleischmann T, Pesaro F, Metzler A (1994) Solar water disinfection: scope of the process and analysis of radiation experiments. *Aqua J Water Supply Res Technol* 43:154–169
155. McGuigan KG, Joyce TM, Conroy RM, Gillespie JB, Elmore-Meegan M (1998) Solar disinfection of drinking water contained in transparent plastic bottles: characterizing the bacterial inactivation process. *J Appl Microbiol* 84:1138–1148
156. Navtoft C, Ubomba-Jaswa E, McGuigan KG, Fernández-Ibáñez P (2008) Effectiveness of solar disinfection using batch reactors with non-imaging aluminium reflectors under real conditions: natural well-water and solar light. *J Photochem Photobiol B* 93:155–161
157. Sommer B, Mariño A, Solarte Y, Salas ML, Dierolf C, Valiente C et al (1997) SODIS – An emerging water treatment process. *J Water Supply Res Technol AQUA* 46:127–137
158. Spiewak I, Benmair R, Messalem R, Radchenko O (1998) Water detoxification and disinfection using high solar concentration and homogeneous photocatalysts. *J Chem Sci* 110:229–238
159. Ubomba-Jaswa E, Fernández-Ibáñez P, Navtoft C, Polo-López MI, McGuigan KG (2010) Investigating the microbial inactivation efficiency of a 25 L batch solar disinfection (SODIS) reactor enhanced with a compound parabolic collector (CPC) for household use. *J Chem Technol Biotechnol* 85:1028–1037
160. Hijnen WAM, Beerendonk EF, Medema GJ (2006) Inactivation credit of UV radiation for viruses, bacteria and protozoan (oo) cysts in water: a review. *Water Res* 40:3–22
161. Nikaido SS, Johnson CH (2000) Daily and circadian variation in survival from ultraviolet radiation in *Chlamydomonas reinhardtii*. *Photochem Photobiol* 71:758–765
162. Nicholson WL, Munakata N, Horneck G, Melosh HJ, Setlow P (2000) Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. *Microbiol Mol Biol Rev* 64:548–572
163. Zepp RG, Erickson DJ III, Paul ND, Sulzberger B (2007) Interactive effects of solar UV radiation and climate change on biogeochemical cycling. *Photochem Photobiol Sci* 6:286–300
164. Zepp RG, Callaghan TV, Erickson DJ (1998) Effects of enhanced solar ultraviolet radiation on biogeochemical cycles. *J Photochem Photobiol B* 46:69–82
165. Zepp RG, Callaghan TV, Erickson DJ (1995) Effects of increased solar ultraviolet radiation on biogeochemical cycles. *Ambio* 24:181–187
166. Sagripanti JL, Lytle CD (2007) Inactivation of influenza virus by solar radiation. *Photochem Photobiol* 83:1278–1282
167. Kripke ML (1984) Immunologic unresponsiveness induced by UV radiation. *Immunol Rev* 80:87–102
168. Lipsitch M, Viboud C (2009) Influenza seasonality: lifting the fog. *Proc Natl Acad Sci USA* 106:3645–3646
169. Hollosy F (2002) Effects of ultraviolet radiation on plant cells. *Micron* 33:179–197
170. Prasad PVV, Kakani VG, Reddy KR (2003) Plants and the environment: ozone depletion. In: *Encyclopedia of applied plant sciences*. Elsevier Academic, Oxford, pp 749–756
171. Jagger J (1967) Introduction to research in photobiology. Prentice Hall, Englewood Cliffs
172. Kakani VG, Reddy KR, Zhao D, Sailaja K (2003) Field crop responses to ultraviolet-B radiation: a review. *Agric For Meteorol* 120:191–218
173. Roosien J, van Klaveren P, Van Vloten-Doting L (1983) Competition between the RNA 3 molecules of wildtype alfalfa mosaic virus and the temperature-sensitive mutant Tb ts 7(UV). *Plant Mol Biol* 2:113–118
174. Cuadra P, Herrera RI, Fajardo V (2004) Effects of UV-B radiation on the Patagonian *Jaborosa magellanica* Brisben. *J Photochem Photobiol B* 76:61–68
175. Mohr H, Schopfer P (1992) *Pflanzenphysiologie*, 4th edn. Springer, Berlin/Heidelberg
176. Mazza CA, Battista D, Zima AM, Szwarcberg BM, Giordano CV, Acevedo A et al (1999) The effects of solar ultraviolet-B radiation on the growth and yield of barley are accompanied by increased DNA damage and antioxidant responses. *Plant Cell Environ* 22:61–70

177. Correia CM, Coutinho JF, Bjorn LO, Torres PJMG (2000) Ultra-violet-B radiation and nitrogen effects on growth and yield of maize under Mediterranean field conditions. *Eur J Agron* 12:117–125
178. Pal M, Sengupta UK, Srivastava AC, Jain V, Meena RC (1999) Changes in growth and photosynthesis of mungbean induced by UV-B radiation. *Indian J Plant Physiol* 4:79–84
179. Mepsted R, Paul N, Stephen J, Nogues S, Corlett JE, Baker NR et al (1996) Effects of enhanced UV-B radiation on pea (*Pisum sativum* L.) grown under field conditions. *Glob Change Biol* 2:325–334
180. Tevini M, Mark U, Saile-Mark M (1991) Effects of enhanced solar UV-B radiation on growth and function of crop plant seedlings. *Curr Top Plant Biochem Physiol* 10:13–31
181. Gonzalez R, Mepsted R, Wellburn AR, Paul ND (1998) Non-photosynthetic mechanisms of growth reduction in pea (*Pisum sativum* L.) exposed to UV-B radiation. *Plant Cell Environ* 21:23–32
182. Kumagai T, Hidema J, Kang HS, Sato T (2001) Effects of supplemental UV-B radiation on the growth and yield of two cultivars of Japanese lowland rice (*Oryza sativa* L.) under the field in a cool rice-growing region of Japan. *Agric Ecosyst Environ* 83:201–208
183. Taylor RM, Tobin AK, Bray CM (1997) DNA damage and repair in plants. In: Lumsden PJ (ed) *Plants and UVB responses to environmental changes*. Cambridge University Press, Cambridge
184. Robberecht R, Caldwell MM (1978) Leaf transmittance of ultraviolet radiation and its implications for plant sensitivity to ultraviolet-radiation induced injury. *Oecologia* 32:277–287
185. Caldwell MM, Robberecht R, Flint SD (1983) Internal filters: prospects for UV-accumulation in higher plants. *Physiol Plant* 58:445–450
186. Tevini M, Iwanzik W, Teramura AH (1983) Effects of UV-B radiation on plants during mild water stress. II Effects on growth, protein, and flavonoid content. *Z Pflanzenphysiol* 110:459–467
187. Beggs CJ, Schneider-Ziebert U, Wellmann E (1986) UV-B radiation and Adaptive Mechanisms in Plants. In: Worrest RC, Caldwell MM (eds) *Stratospheric ozone reduction, solar ultraviolet radiation and plant life*. Springer, Berlin, pp 235–250
188. Bischof K, Wiencke C (2006) Auswirkung der Zunahme der UV-Strahlung. In: Lozan J, Graál H, Hubberten HW, Hupfer P, Piepenburg D (eds) *Warnsignale aus den Polarregionen*. Druckerei in St. Pauli, Hamburg, pp 259–263
189. Wiencke C, Roleda MY, Gruber A, Clayton MN, Bischof K (2006) Susceptibility of zoospores to UV radiation determines upper depth distribution limit of Arctic kelps: evidence through field experiments. *J Ecol* 94:455–463
190. Johansson G, Snoeijjs P (2002) Macroalgal photosynthetic responses to light in relation to thallus morphology and depth zonation. *Mar Ecol Prog Ser* 244:63–72
191. Roleda MY, Wiencke C, Hanelt D, van de Poll WH, Gruber A (2005) Sensitivity of Laminariales zoospores from Helgoland (North Sea) to ultraviolet and photosynthetically active radiation: implications for depth distribution and seasonal reproduction. *Plant Cell Environ* 28:466–479
192. Lotze HK, Worm B, Molis M, Wahl M (2002) Effects of UV radiation and consumers on recruitment and succession of a marine macrobenthic community. *Mar Ecol Prog Ser* 243:57–66
193. Bischof K, Hanelt D, Aguilera J, Karsten U, Vögele B, Sawall T et al (2002) Seasonal variation in ecophysiological patterns in macroalgae from an arctic fjord. I. Sensitivity of photosynthesis to ultraviolet radiation. *Mar Biol* 140:1097–1106
194. de Bakker NVJ, van Bodegom PM, van de Poll WH, Boelen E, Nat E, Rozema J et al (2005) Is UV-B radiation affecting charophycean algae in shallow freshwater systems? *New Phytol* 166:957–966
195. Martinez-Abaigar J, Nunez-Olivera E, Beaucourt N, Garcia-Alvaro MA, Tomas R, Arroniz M (2003) Different physiological responses of two aquatic bryophytes to enhanced ultraviolet-B radiation. *J Bryol* 25:17–30
196. Figueroa FL, Jimenez C, Vinegla B, Perez-Rodriguez E, Aguilera J, Flores-Moya A et al (2002) Effects of solar UV radiation on photosynthesis of the marine angiosperm *Posidonia oceanica* from southern Spain. *Mar Ecol Prog Ser* 230:59–70
197. Brandt LA, Koch EW (2003) Periphyton as a UV-B filter on seagrass leaves: a result of different transmittance in the UV-B and PAR ranges. *Aquat Bot* 76:317–327
198. Li FR, Peng SL, Chen BM, Hou YP (2010) A meta-analysis of the responses of woody and herbaceous plants to elevated ultraviolet-B radiation. *Acta Oecologica* 36:1–9
199. Rozema J, Van de Staaij JWM, Tosserams M (1997) Effects of UV-B radiation on plants from agro- and natural ecosystems. In: Lumsden PJ (ed) *Plants and UV-B responses to environmental change*. Cambridge University Press, Cambridge
200. Caldwell MM, Flint SD (1994) Solar ultraviolet radiation and ozone layer change: implications for crop plants. In: ASo A (ed) *Physiology and determination of crop yield*. American Society of Agronomy, Madison
201. Sullivan JH, Teramura AH (1992) The effects of UV-B radiation on loblolly pine. 2. Growth of field-grown seedlings. *Tree Struct Funct* 6:115–120
202. Naidu SL, Sullivan JH, Teramura AH, DeLucia EH (1993) The effects of ultraviolet-B radiation on photosynthesis of different aged needles in field-grown loblolly-pine. *Tree Physiol* 12:151–162
203. Bassman JH, Edwards GE, Robberecht R (2002) Long-term exposure to enhanced UV-B radiation is not detrimental to growth and photosynthesis in Douglas-fir. *New Phytol* 154:107–120
204. Trost-Sedej T, Gaberscik A (2008) The effects of enhanced UV-B radiation on physiological activity and growth of Norway spruce planted outdoors over 5 years. *Trees* 22:423–435
205. McLeod AR, Newsham KK (1997) Impacts of elevated UV-B on forest ecosystems. In: Lumsden PJ (ed) *Plants and UV-B responses to environmental change*, 64th edn. Cambridge University Press, Cambridge

206. Karabourniotis G, Papadopoulos K, Papamarkou M, Manetas Y (1992) Ultraviolet-B radiation absorbing capacity of leaf hairs. *Physiol Plant* 86:414–418
207. Karabourniotis G, Kyprisiss A, Manetas Y (1993) Leaf hairs of *Olea europaea* protect underlying tissues against ultraviolet-B radiation damage. *Environ Exp Bot* 33:341–345
208. Grammatikopoulos G, Karabourniotis G, Kyprisiss A, Petropoulou Y, Manetas Y (1994) Leaf hairs of olive (*Olea europaea*) prevent stomatal closure by ultraviolet-B radiation. *Aust J Plant Physiol* 21:293–301
209. Day TA, Howells BW, Rice WJ (1994) Ultraviolet absorption and epidermal-transmittance spectra in foliage. *Physiol Plant* 92:207–218
210. Nedunchezian N, Kulandaivelu GL (1991) Effects of UV-B enhanced radiation on ribulose 1, 5-bisphosphate carboxylase in leaves of *Vigna sinensis*. *Photosynthetica* 25:431–435
211. Premkumar A, Kulandaivelu G (1996) Influence of ultraviolet-B enhanced solar radiation on growth and photosynthesis of potassium deficient cowpea seedlings. *Photosynthetica* 32:521–528
212. Basiouny FM (1986) Sensitivity of corn, oats, peanuts, rice, rye, sorghum, soybean and tobacco to UV-B radiation under growth chamber conditions. *J Agron Crop Sci* 157:31–35
213. Pal M, Jain V, Sengupta UK (1998) Influence of enhanced UV-B radiation on mustard: cultivar response. *Indian J Plant Physiol* 3:188–193
214. Singh A (1995) Influence of enhanced UV-B radiation on tropical legumes. *Trop Ecol* 36:249–252
215. Li Y, Yue M, Wang XL (1998) Effects of enhanced ultraviolet-B radiation on crop structure, growth and yield components of spring wheat under field conditions. *Field Crops Res* 57:253–263
216. Teramura AH, Murali NS (1986) Intraspecific differences in growth and yield of soybean exposed to ultraviolet-B radiation under greenhouse and field conditions. *Environ Exp Bot* 26:89–95
217. Al-Oudat M, Baydoun SA, Mohammad A (1989) Effects of enhanced UV-B on growth and yield of two Syrian crops wheat (*Triticum durum* var. Horani) and broad beans (*Vicia faba*) under field conditions. *Environ Exp Bot* 40(1):11–16
218. Finnie JW, Bostock DE (1979) Skin neoplasia in dogs. *Aust Vet J* 55:602–604
219. Fairham J, Harcourt-Brown FM (1999) Preliminary investigation of the vitamin D status of pet rabbits. *Vet Rec* 145: 452–454
220. Hazewinkel HAW, Tryfonidou MA (2002) Vitamin D₃ metabolism in dogs. *Mol Cell Endocrinol* 197:23–33
221. Morris JG (1999) Ineffective vitamin D synthesis in cats is reversed by an inhibitor of 7-dehydrocholesterol- Δ^7 -reductase. *J Nutr* 129:903–908
222. How KL, Hazewinkel HAW, Mol JA (1994) Dietary vitamin D dependence of cat and dog due to inadequate cutaneous synthesis of vitamin D. *Gen Comp Endocrinol* 96:12–18
223. How KL, Hazewinkel HA, Mol JA (1995) Photosynthesis of vitamin D in the skin of dogs cats and rats. *Vet Q* 17(Suppl 1):S29
224. Brommage R, Miller SC, Langman CB, Bouillon R, Smith R, Bourdeau JE (1988) The effects of chronic vitamin D deficiency on the skeleton in the adult rabbit. *Bone* 9:131–139
225. Mellanby RJ, Mee AP, Berry JL, Herrtage ME (2005) Hypercalcaemia in two dogs caused by excessive dietary supplementation of vitamin D. *J Small Anim Pract* 46: 334–338
226. Schulze C, Rothuizen J, Van Sluijs FJ, Hazewinkel HAW, Van Den Ingh TSGAM (2000) Extrahepatic biliary atresia in a border collie. *J Small Anim Pract* 41:27–30
227. Ladds PW, Kraft H, Sokale A, Trueman KF (1983) Neoplasms of the skin of dogs in tropical Queensland. *Aust Vet J* 60:87–88
228. Teifke JP, Lohr CV (1996) Immunohistochemical detection of P53 overexpression in paraffin wax-embedded squamous cell carcinomas of cattle, horses, cats and dogs. *J Comp Pathol* 114:205–210
229. UNEP (1991) Environmental effects of ozone depletion: 1991 update. United Nations Environment Programme (UNEP), Nairobi
230. Gross TL, Ihrke PJ, Walder EJ, Affolter VK (2005) Skin diseases of the dog and cat: clinical and histopathologic diagnosis. Blackwell, Oxford
231. Moss LC, Severin GA (1963) Tattooing as therapy for chronic solar dermatitis (Collie nose) in the dog. *J Am Vet Med Assoc* 142:609–610
232. Nikula KJ, Benjamin SA, Angleton GM, Saunders WJ, Lee AC (1992) Ultraviolet radiation, solar dermatosis, and cutaneous neoplasia in beagle dogs. *Radiat Res* 129:11–18
233. Nielsen SW, Cole CR (1960) Cutaneous epithelial neoplasms of the dog—a report of 153 cases. *Am J Vet Res* 21:931
234. Nesbitt GH (1983) Canine and feline dermatology: a systematic approach. Lea & Febiger, Philadelphia
235. Hargis AM, Thomassen RW, Phemister RD (1977) Chronic dermatosis and cutaneous squamous cell carcinoma in the beagle dog. *Vet Pathol* 14:218–228
236. Jörgen K (1988) Skin tumors in cats. Occurrence and frequency in the research material (biopsies from 1984 to 1987) of the Institute for Veterinary Pathology, Zurich. *Schweiz Arch Tierheilk* 130:559–569
237. Dorn CR, Taylor DO, Schneider R (1971) Sunlight exposure and risk of developing cutaneous and oral squamous cell carcinomas in white cats. *J Natl Cancer Inst* 46:1073–1078
238. Almeida EMP, Caraca RA, Adam RL, Souza EM, Metze K, Cintra ML (2008) Photodamage in feline skin: clinical and histomorphometric analysis. *Vet Pathol* 45:327–335
239. Slatyer DH, Edwards ME, Hawkins CD, Wilcox GE (1982) A national survey of the occurrence of infectious bovine keratoconjunctivitis. *Aust Vet J* 59:65–68
240. Méndez A, Pérez J, Ruiz-Villamor E, García R, Martín MP, Mozos E (1997) Clinicopathological study of an outbreak of squamous cell carcinoma in sheep. *Vet Rec* 141:597–600

241. Payne RJ, Lean MS, Greet TRC (2009) Third eyelid resection as a treatment for suspected squamous cell carcinoma in 24 horses. *Vet Rec* 165:740–743
242. Heeney JL, Valli VEO (1985) Bovine ocular squamous cell carcinoma: an epidemiological perspective. *Can J Comp Med* 49:21–26
243. Sloss V, Smith TJ, De Yi G (1986) Controlling ocular squamous cell carcinoma in Hereford cattle. *Aust Vet J* 63:248–251
244. Russell WC, Brinks JS, Kainer RA (1976) Incidence and heritability of ocular squamous cell tumors in Hereford cattle. *J Anim Sci* 43:1156–1162
245. Anderson DE, Badzioch M (1991) Association between solar radiation and ocular squamous cell carcinoma in cattle. *Am J Vet Res* 52:784–788
246. Gharagozlu MJ, Hekmati P, Ashrafihelan J (2007) A clinical and histopathological study of ocular neoplasms in dairy cattle. *Veterinarski Arh* 77:409–426
247. Morris CA (2007) A review of genetic resistance to disease in *Bos Taurus* cattle. *Vet J* 174:481–491
248. Anderson DE, Skinner PE (1961) Studies on bovine ocular squamous carcinoma ("Cancer Eye") XI. Effects of sunlight. *J Anim Sci* 20:474–477
249. Radostits OM, Arundel JH, Gay CC, Blood DC, Hinchcliff KW (2000) *Veterinary medicine: a textbook of the diseases of cattle, sheep, pigs, goats and horses*. W.B. Saunders, London
250. Woodward RR, Knapp B Jr (1950) The hereditary aspect of eye cancer in hereford cattle. *J Anim Sci* 9:578–581
251. Blackwell RL, Anderson DE, Knox JH (1956) Age incidence and heritability of cancer eye in hereford cattle. *J Anim Sci* 15:943–951
252. SGdS W (1974) Das Vulvacarcinom beim Rind auf Ceylon. *Zbl Vet Med A21*:834–843
253. Hühnermund G (1973) Disposition der Ayreshirerasse zu karzinomatösen Veränderungen an Vulva und Membrana nictans bei extremen Sonneneinwirkungen in den Tropen. *Berl Münchn Tierärztl Wsch* 21:414–415
254. Yeruham I, Perl S, Orgad U, Yakobson B (1999) Tumours of the vulva and vagina in cattle – a 10-year survey. *Vet J* 158: 237–239
255. Wettimuny SG (1974) Bovine vulval carcinoma in Ceylon. *Zentralbl Veterinärmed A* 21:834
256. Kopecky KE, Pugh GW Jr, Hughes DE (1980) Wavelength of ultraviolet radiation that enhances onset of clinical infectious bovine keratoconjunctivitis. *Am J Vet Res* 41:1412–1415
257. Forrest JW, Fleet MR (1985) Lack of tanning in white Merino sheep following exposure to ultraviolet light. *Aust Vet J* 62:244–246
258. Turner AS, Mansmann RA, McAllister ES, Pratt PW (1982) *Equine medicine and surgery*, 3rd edn. American Veterinary, Santa Barbara
259. Dugan SJ, Curtis CR, Roberts SM, Severin GA (1991) Epidemiologic study of ocular/adnexal squamous cell carcinoma in horses. *J Am Vet Med Assoc* 198:251–256
260. Eberlein B, Bergner T, Przybilla B (1992) Demonstration of olaquinox phototoxicity in vitro. *Photodermatol Photoimmunol Photomed* 9:63–66
261. Bovenschen HJ, Peters B, Koetsier MI, Van Der Valk PG (2009) Occupational contact dermatitis due to multiple sensitizations in a pig farmer. *Contact Dermat* 61:127–128
262. Schauder S, Schröder W, Geier J (1996) Olaquinox-induced airborne photoallergic contact dermatitis followed by transient or persistent light reactions in 15 pig breeders. *Contact Dermat* 35:344–354
263. Belhadjali H, Marguery MC, Journé F, Giordano-Labadie F, Lefebvre H, Bazex J (2002) Allergic and photoallergic contact dermatitis to Olaquinox in a pig breeder with prolonged photosensitivity. *Photodermatol Photoimmunol Photomed* 18:52–53
264. Bartussek H (1988) Haltung. In: Haiger A, Storhas R, Bartussek H (eds) *Artgemäße Nutztierhaltung*. Ulmer, Stuttgart
265. Sozonov JI, Inesin VA (1975) Ergebnisse der UV-Bestrahlung von Kälbern. *Landwirtschaftliches Zentralbl* 4:460
266. Hidioglou M, Proulx JG, Roubos D (1979) 25-hydroxyvitamin D in plasma of cattle. *J Dairy Sci* 62:1076–1080
267. Horst RL, Littledike ET, Riley JL, Napoli JL (1981) Quantitation of vitamin D and its metabolites and their plasma concentrations in five species of animals. *Anal Biochem* 116:189–203
268. Hidioglou M, Williams CJ, Proulx JG (1985) Plasma vitamin D3 response in cattle and sheep exposed to ultraviolet radiation. *Int J Vitam Nutr Res* 55:41–46
269. Flachowsky E, Graf H, Matthey M, Ochrimenko WI, Beyersdorfer S, Dorn W et al (1993) Einfluß von Jahreszeit, Haltungsform und einer Vit.D3 Zulage auf den Vit.A, D und E Gehalt von Kuhmilch sowie die 25-OH Vit-D3-Konzentration im Blutplasma von Milchkühen. *Mh Vetmed* 48:197–202
270. Richter GH, Flachowsky G, Matthey M, Ochrimenko K, Wolfram D, Schade T (1990) Einfluß der Stall- bzw. Auslaufhaltung auf die 25-OH-Vitamin-D3-Blutplasmakonzentration bei Mastbullen. *Mh Vet Med* 45:227
271. Hidioglou M (1987) Kinetics of intravenously administered 25-hydroxyvitamin D3 in sheep and the effect of exposure to ultraviolet radiation. *J Anim Sci* 65:808–814
272. Hidioglou M, Karpinski K (1989) Providing vitamin D to confined sheep by oral supplementation vs ultraviolet irradiation. *J Anim Sci* 67:794–802
273. Smith BS, Wright H (1984) Relative contributions of diet and sunshine to the overall vitamin D status of the grazing ewe. *Vet Rec* 115:537–538
274. Zintzen H, Boyazoglu PA (1973) Vitamin D and its relation to calcium phosphorus metabolism in ruminants. *J S Afr Vet Med Assoc* 44:25–36
275. El Shorafa WM, Feaster JP, Ott EA, Asquith RL (1979) Effect of vitamin D and sunlight on growth and bone development of young ponies. *J Anim Sci* 48:882–886
276. Piccione G, Assenza A, Fazio F, Bergero D, Caola G (2008) Daily rhythms of serum vitamin D-metabolites, calcium and phosphorus in horses. *Acta Vet Brno* 77:151–157

277. Piccione G, Assenza A, Grasso F, Caola G (2004) Daily rhythm of circulating fat soluble vitamin concentration (A, D, E and K) in the horse. *J Circadian Rhythms* 2(1):3
278. Hart EB, Steenbock H, Lepkovsky S, Halpin JG (1923) The nutritional requirements of baby chicks: III. The relation of light to the growth of the chicken. *J Biol Chem* 58:33–42
279. Bethke RM, Kennard DC, Kik MC (1925) Nutritional studies of the growing chick: I. The relation of sunlight and green clover to leg weakness in chicks. *J Biol Chem* 63:377–390
280. Mussehl FE, Bancroft PM (1925) Nutrient requirements of growing chicks. *Poult Sci* 4:118–122
281. Heuser GF, Norris LC (1929) Rickets in chicks: III. The effectiveness of mid-summer sunshine and irradiation from quartz mercury vapour arc in preventing rickets. *Poult Sci* 8:89–98
282. Scott ML, Nesheim MC, Young RJ (1982) Nutrition of the chicken. M. L. Scott and Associates, Ithaca
283. Bernard JB, Watkins BE, Ullrey DE (1989) Manifestations of vitamin D deficiency in chicks reared under different artificial lighting regimes. *Zoo Biol* 8:313–395
284. Ferguson GW, Jones JR, Gehrmann WH, Hammack SH, Talent LG, Hudson RD et al (1996) Indoor husbandry of the panther chameleon *Chamaeleo [Furcifer] pardalis*: effects of dietary vitamins A and D and ultraviolet irradiation on pathology and life-history traits. *Zoo Biol* 15:279–299
285. Ferguson GW, Gehrmann WH, Chen TC, Dierenfeld ES, Holick MF (2002) Effects of artificial ultraviolet light exposure on reproductive success of the female panther chameleon (*Furcifer pardalis*) in captivity. *Zoo Biol* 21:525–537
286. Narbaitz R, Tsang CPW (1989) Vitamin D deficiency in the chick embryo: effects on prehatching motility and on the growth and differentiation of bones, muscles, and parathyroid glands. *Calcif Tissue Int* 44:348–355
287. Packard MJ, Clark NB (1996) Aspects of calcium regulation in embryonic lepidosaurs and chelonians and a review of calcium regulation in embryonic archosaurs. *Physiol Zool* 69:435–466
288. Häder DP, Sinha RP (2005) Solar ultraviolet radiation-induced DNA damage in aquatic organisms: Potential environmental impact. *Mutat Res Fundam Mol Mech Mutagen* 571:221–233
289. Madronich S, Flocke S (1995) Theoretical estimation of biologically effective UV radiation at the Earth's surface. In: Zerefos CS, Bais AF (eds) *Solar ultraviolet radiation: modeling, measurements, and effects*, pp 23–48
290. Rozema J, Björn LO, Bornman JF, Gaberscik A, Häder DP, Trost T et al (2002) The role of UV-B radiation in aquatic and terrestrial ecosystems-An experimental and functional analysis of the evolution of UV-absorbing compounds. *J Photochem Photobiol B* 66:2–12
291. Huot Y, Jeffrey WH, Davis RF, Cullen JJ (2000) Damage to DNA in bacterioplankton: a model of damage by ultraviolet radiation and its repair as influenced by vertical mixing. *Photochem Photobiol* 72:62–74
292. Zepp RG (2003) UV exposure of coral assemblages in the Florida Keys. Environmental Protection Agency Report No. 600/R-03/095. U.S. Environmental Protection Agency, Washington, DC
293. Kouwenberg JHM, Lantoine F (2007) Effects of ultraviolet-B stressed diatom food on the reproductive output in Mediterranean *Calanus helgolandicus* (Crustacea; Copepoda). *J Exp Mar Biol Ecol* 341:239–253
294. Kouwenberg JHM, Browman HI, Runge JA, Cullen JJ, Davis RF, St-Pierre JF (1999) Biological weighting of ultraviolet (280–400 nm) induced mortality in marine zooplankton and fish. II. *Calanus finmarchicus* (Copepoda) eggs. *Mar Biol* 134:285–293
295. Hunter JR, Taylor JH, Moser HG (1979) Effect of ultraviolet irradiation on eggs and larvae on the northern anchovy, *Engraulis mordax*, and the Pacific mackerel, *Scomber japonicus*, during the embryonic stage. *Photochem Photobiol* 29:325–338
296. Bell G, Hoar W (1950) Some effects of ultraviolet radiation on sockeye salmon eggs and alevins. *Can J Res* 28:35–43
297. Zagarese HE, Williamson CE (2001) The implications of solar UV radiation exposure for fish and fisheries. *Fish Fish* 2: 250–260
298. Kouwenberg JHM, Browman HI, St-Pierre JF, Runge JA, Cullen JJ, Davis RF (1999) Biological weighting of ultraviolet (280–400 nm) induced mortality in marine zooplankton and fish. I. Atlantic cod (*Gadus morhua*) eggs. *Mar Biol* 134:269–284
299. Brocklebank JR, Armstrong RD (1994) Solar dermatitis in hatchery-reared salmonids in British Columbia. *Can Vet J* 35:651–652
300. Lowe C, Goodman-Lowe G (1996) Suntanning in hammerhead sharks. *Nature* 383:677
301. Dunbar C (1959) Sunburn in fingerling rainbow trout. *Prog Fish Cult* 21:74
302. Bullock AM, Coutts RR (1985) The impact of solar ultraviolet radiation upon the skin of rainbow trout, *Salmo gairdneri* Richardson, farmed at high altitude in Bolivia. *J Fish Dis* 8:263–272
303. McArdle J, Bullock AM (1987) Solar ultraviolet radiation as a causal factor of "summer syndrome" in cage-reared Atlantic salmon, *Salmo solar* L.: a clinical and histopathological study. *J Fish Dis* 10:255–264
304. Kaweewat K, Hofer R (1997) Effect of UV-B radiation on goblet cells in the skin of different fish species. *J Photochem Photobiol B* 41:222–226
305. Rodger HD (1991) Summer lesion syndrome in salmon: a retrospective study. *Vet Rec* 129:237–239
306. Bullock AM, Roberts RJ, Waddington P, Bookless WD (1983) Sunburn lesions in koi carp. *Vet Rec* 112:551
307. Markkula SE, Karvonen A, Salo H, Valtonen ET, Jokinen EI (2007) Ultraviolet B irradiation affects resistance of rainbow trout (*Oncorhynchus mykiss*) against bacterium *Yersinia ruckeri* and trematode *Diplostomum spathaceum*. *Photochem Photobiol* 83:1263–1269
308. Roberts R (2001) *Fish pathology*. Churchill Livingstone, Edinburgh

309. Allison L (1960) "Sunburning" fingerling lake trout with ultraviolet light and the effect off a niacin-fortified diet. *Prog Fish Cult* 22:114–116
310. Wall AE (1998) Cataracts in farmed Atlantic salmon (*Salmo salar*) in Ireland, Norway and Scotland from 1995 to 1997. *Vet Rec* 142:626–630
311. Karentz D, McEuen FS, Land MC, Dunlap WC (1991) Survey of mycosporine-like amino acid compounds in Antarctic marine organisms: Potential protection from ultraviolet exposure. *Mar Biol* 108:157–166
312. Uitz J, Claustre H, Morel A, Hooker SB (2006) Vertical distribution of phytoplankton communities in open ocean: an assessment based on surface chlorophyll. *J Geophys Res* 111: C08005
313. Kessler K, Lockwood RS, Williamson CE, Saros JE (2008) Vertical distribution of zooplankton in subalpine and alpine lakes: Ultraviolet radiation, fish predation, and the transparency-gradient hypothesis. *Limnol Oceanogr* 53:2374–2382
314. Yamashita Y, Tanoue E (2009) Basin scale distribution of chromophoric dissolved organic matter in the Pacific Ocean. *Limnol Oceanogr* 54:598–609
315. Loiselle SA, Bracchini L, Dattilo AM, Ricci M, Tognazzi A, Cózar A et al (2009) Optical characterization of chromophoric dissolved organic matter using wavelength distribution of absorption spectral slopes. *Limnol Oceanogr* 54:590–597
316. Browman HI, Vetter RD, Rodriguez CA, Cullen JJ, Davis RF, Lynn E et al (2003) Ultraviolet (280–400 nm)-induced DNA damage in the eggs and larvae of *Calanus finmarchicus* G. (Copepoda) and Atlantic cod (*Gadus morhua*). *Photochem Photobiol* 77:397–404
317. Bullock AM (1982) The pathological effects of ultraviolet radiation on the epidermis of teleost fish with reference to the solar radiation effect in higher animals. *Proc R Soc Edinb* 81B:199–210
318. Bullock AM, Roberts RJ (1979) Skin photosensitisation by phenothiazine in cultured rainbow trout (*Salmo gairdneri* Richardson). *Vet Rec* 104:55
319. Bullock AM, Roberts RJ (1981) Sunburn lesions in salmonid fry: a clinical and histological report. *J Fish Dis* 4:271–275
320. Bullock AM (1985) The effect of ultraviolet-B radiation upon the skin of the plaice, *Pleuronectes platessa* L., infected with the bodonid ectoparasite *Ichthyobodo necator* (Henneguy, 1883). *J Fish Dis* 8:547–550
321. Jokinen IE, Markkula ES, Salo HM, Kuhn P, Nikoskelainen S, Arts MT et al (2008) Exposure to increased ambient ultraviolet B radiation has negative effects on growth, condition and immune function of juvenile Atlantic salmon (*Salmo salar*). *Photochem Photobiol* 84:1265–1271
322. Markkula SE, Salo HM, Immonen AK, Jokinen EI (2005) Effects of short- and long-term ultraviolet B irradiation on the immune system of the common carp (*Cyprinus carpio*). *Photochem Photobiol* 81:595–602
323. Armstrong TN, Reimschuessel R, Bradley BP (2002) DNA damage, histological changes and DNA repair in larval Japanese medaka (*Oryzias latipes*) exposed to ultraviolet-B radiation. *Aquat Toxicol* 58:1–14
324. Alemanni ME, Lozada M, Zagarese HE (2003) Assessing sub-lethal effects of ultraviolet radiation in juvenile rainbow trout (*Oncorhynchus mykiss*). *Photochem Photobiol Sci* 2:867–870
325. Bullock AM (1982) The pathological effects of ultraviolet radiation on the epidermis of teleost fish with reference to the solar radiation effect in higher animals. *Proc R Soc Edinb B Biol Sci* 81:199–210
326. Nowak BF (1999) Significance of environmental factors in aetiology of skin diseases of teleost fish. *Bull Eur Assoc Fish Pathol* 19:290–292
327. Handlinger J, Soltani M, Percival S (1997) The pathology of *Flexibacter maritimus* in aquaculture species in Tasmania, Australia. *J Fish Dis* 20:159–168
328. Cullen AP, Monteith-McMaster CA, Sivak JG (1994) Lenticular changes in rainbow trout following chronic exposure to UV radiation. *Curr Eye Res* 13:731–737
329. Nelson PA, Kajiura SM, Losey GS (2003) Exposure to solar radiation may increase ocular UV-filtering in the juvenile scalloped hammerhead shark, *Sphyrna lewini*. *Mar Biol* 142:53–56
330. Zeng Z, Richardson J, Verduzco D, Mitchell DL, Patton EE (2009) Zebrafish have a competent p53-dependent nucleotide excision repair pathway to resolve ultraviolet B-induced DNA damage in the skin. *Zebrafish* 6:405–415
331. Banaszak AT (2003) Photoprotective physiological and biochemical responses of aquatic organisms. In: Helbling EW, Zagarese HE, Zagarese H (eds) *UV effects in aquatic organisms and ecosystems*. Royal Society of Chemistry, Cambridge, UK, pp 329–356
332. Eckes MJ, Siebeck UE, Dove S, Grutter AS (2008) Ultraviolet sunscreens in reef fish mucus. *Mar Ecol Prog Ser* 353:203–211
333. Zamzow JP (2007) Ultraviolet-absorbing compounds in the mucus of shallow-dwelling tropical reef fishes correlate with environmental water clarity. *Mar Ecol Prog Ser* 343:263–271
334. Zamzow JP (2004) Effects of diet, ultraviolet exposure, and gender on the ultraviolet absorbance of fish mucus and ocular structures. *Mar Biol* 144:1057–1064
335. Rick IP, Bakker TCM (2008) Color signaling in conspicuous red sticklebacks: do ultraviolet signals surpass others? *BMC Evol Biol* 8:189
336. Gacic Z, Damjanovic I, Mickovic B, Hegedis A, Nikcevic M (2007) Spectral sensitivity of the dogfish shark (*Scyliorhinus canicula*). *Fish Physiol Biochem* 33:21–27
337. Matsumoto T, Ihara H, Ishida Y, Yamamoto S, Murata O, Ishibashi Y (2010) Spectral sensitivity of juvenile chub mackerel (*Scomber japonicus*) in visible and ultraviolet light. *Fish Physiol Biochem* 36(1):63–70
338. Leech DM, Johnsen S (2003) Behavioral response – UVR avoidance and vision. In: Helbling EW, Zagarese HE, Zagarese H (eds) *UV effects in aquatic organisms and ecosystems*. Royal Society of Chemistry, London, pp 329–356

339. Leech DM, Johnsen S (2009) Light, biological receptors. In: Likens GE (ed) *Encyclopedia of inland waters*. Elsevier, Oxford, pp 671–681
340. Holtby LB, Bothwell ML (2008) Effects of solar ultraviolet radiation on the behaviour of juvenile coho salmon (*Oncorhynchus kisutch*): avoidance, feeding, and agonistic interactions. *Can J Fish Aquat Sci* 65:701–711
341. Rao DS, Raghuramulu N (1996) Food chain as origin of vitamin D in fish. *Comp Biochem Physiol A Physiol* 114:15–19
342. Copping AM (1934) Origin of vitamin D in cod-liver oil: vitamin D content of zooplankton. *Biochem J* 28:1516–1520
343. Rao DS, Raghuramulu N (1997) Vitamin D₃ in *Tilapia mossambica*: relevance of photochemical synthesis. *J Nutr Sci Vitaminol* 43:425–433
344. Lock EJ, Waagbo R, Wendelaar Bonga S, Flik G (2010) The significance of vitamin D for fish: a review. *Aquac Nutr* 16:100–116
345. Ferguson GW, Brinker AM, Gehrmann WH, Bucklin SE, Baines FM, Mackin SJ (2010) Voluntary exposure of some western-hemisphere snake and lizard species to ultraviolet-B radiation in the field: how much ultraviolet-B should a lizard or snake receive in captivity? *Zoo Biol* 29(3):317–334
346. Göhlthenth R (1982) Höhensonne für Zootiere im Zoo Berlin. *Berliner Münchner Tierärztliche Wochenschrift* 95(11):220
347. Morrissey JK, Reichard T, Lloyd M, Bernard J (1995) Vitamin-D-deficiency rickets in three colobus monkeys (*Colobus guereza kikuyuensis*) at the Toledo Zoo. *J Zoo Wildl Med* 26(4):564–568
348. Hofmann-Parisot M (1995) Result of a European-wide inquiry about the use of artificial UV radiation sources in zoo animal husbandry with special regard to the mammals (Primates/Simiae/monkeys). *Berl Münch Tierärztl Wschr* 108:185–191
349. Tonge S (1985) The management of juvenile telfair's skinks *leiopisma telfairii* with particular reference to the role of ultra violet light. In: *Reptiles: breeding, behaviour and veterinary aspects*. British Herpetological Society, London, pp 61–72
350. Bosch H, Frank W (1983) Häufige Erkrankungen bei im Terrarium gehaltenen Amphibien und Reptilien. *Salamandra* 19:29–54
351. McCrystal HG, Behler JL (1982) Husbandry and reproduction of captive Giant ameiva lizards. *Int Zoo Yearb* 22:159–163
352. Townsend CR, Cole CJ (1985) Additional notes on requirements of Captive whiptail lizards (*Cnemidophorus*) with emphasis on ultraviolet radiation. *Zoo Biol* 4:49–55
353. Acierno MJ, Mitchell MA, Roundtree MK, Zachariah TT (2006) Effects of ultraviolet radiation on 25-hydroxyvitamin D₃ synthesis in red-eared slider turtles (*Trachemys scripta elegans*). *Am J Vet Res* 67:2046–2049
354. Acierno MJ, Mitchell MA, Zachariah TT, Roundtree MK, Kirchgessner MS, Guzman DSM (2008) Effects of ultraviolet radiation on plasma 25-hydroxyvitamin D₃ concentrations in corn snakes (*Elaphe guttata*). *Am J Vet Res* 69:294–297
355. Gehrmann WH (1987) Ultraviolet irradiances of various lamps used in animal husbandry. *Zoo Biol* 6:117–127
356. Burger RM, Gehrmann WH, Ferguson GW (2007) Evaluation of UVB reduction by materials commonly used in reptile husbandry. *Zoo Biol* 26:417–423
357. Gardiner DW, Baines FM, Pandher K (2009) Photodermatitis and photokeratoconjunctivitis in a ball python (*Python regius*) and a blue-tongue skink (*Tiliqua spp.*). *J Zoo Wildl Med* 40:757–766
358. Farman JC, Gardiner BG, Shanklin JD (1985) Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction. *Nature* 315:207–210
359. Blumthaler M, Ambach W (1990) Indication of increasing solar ultraviolet-B radiation flux in alpine regions. *Science* 248:206–208
360. Rütger TM (2007) How different wavelengths of the ultraviolet spectrum contribute to skin carcinogenesis: The role of cellular damage responses. *J Invest Dermatol* 127: 2103–2105

Books and Reviews

- Björn LO (2008) *Photobiology. The science of life and light*. Springer, New York
- Cockell C, Blaustein AR (2001) *Ecosystems, evolution, and ultraviolet radiation*. Springer, New York
- Gao W, Schmoldt DL, Slusser JR (2010) *UV radiation in global climate change. Measurements, modeling and effects on ecosystems*. Springer, Heidelberg
- Ghetti F, Checcucci G, Bornman JF (2006) *Environmental UV radiation: impact on ecosystems and human health, NATO science series, IV earth and environment sciences*. Springer, Dordrecht
- Helbling EW, Zagarese HE, Zagarese H (2003) *UV effects in aquatic organisms and ecosystems, Comprehensive series in photochemical and photobiological sciences*. Royal Society of Chemistry, Cambridge
- Lumsden PJ (1997) *Plants and UV-B: responses to environmental change, Society for experimental biology seminar series*. Cambridge University Press, Cambridge
- Mora SJD, Demers S, Vernet M (2000) *The effects of UV radiation in the marine environment, Cambridge environment chemistry series*. Cambridge University Press, Cambridge
- Reichrath J (2008) *Sunlight, vitamin D and skin cancer, Advances in experimental medicine and biology*. Springer, New York

